Navigation, findability and the usage of cultural heritage on the web

an exploratory study

Fransson, Jonas

Publication date:
2014

Document version
Early version, also known as pre-print

Navigation, findability and the usage of cultural heritage on the web: an exploratory study

JONAS FRANSSON

Navigation, findability and the usage of cultural heritage on the web: an exploratory study

JONAS FRANSSON

PhD thesis from Royal School of Library and Information Science, Denmark

CIP – Cataloguing in Publication

Fransson, Jonas

Navigation, findability and the usage of cultural heritage on the web: an exploratory study / Jonas Fransson. – Copenhagen: Royal School of Library and Information Science, Copenhagen University, 2014. 284 p.

Available online:
http://forskning.ku.dk/search/publicationdetail/?id=190156da-e8cc-4a10-a75b-10b2da2807e6

ISBN 978-87-7415-325-2

ISBN 978-87-7415-325-2

Navigation, hittbarhet och användningen av kulturarv på webben: en explorativ studie

Jonas Fransson

Ph.d.-afhandling fra Det Informationsvidenskabelige Akademi

Acknowledgments

The process leading to this thesis started during my studies in Library and Information Science at Lund University, in 2002. I got hold of a copy of "The laws of the web" written by Bernando A. Huberman (2001). His patterns of behaviour on an aggregate level fascinated me, and it still does. Later in 2009 at the Royal School of Library and Information Science (RSLIS), University of Copenhagen, I was happy to be able to proceed with a doctoral project on users' web behaviour and findability in a cultural context.

First of all I want to thank my supervisors Niels Ole Pors (1949-2013), Birger Larsen and Peter Ingwersen. Thank you for "sheparding" me through the whole process. I have felt safe, but never controlled or directed. Sadly Niels Ole Pors suddenly passed away last year. His thoughts continue to inspire me. Thank you for everything!

Secondly I want to send my gratitude to the staff at the Royal School of Library and Information Science who has supported me in numerous ways during these years. My gratitude also goes to teachers and fellow students at PhD courses, seniors and participants at doctoral forums as well as in informal sessions and discussions. In addition, especially to my "buddy" Lennart Björneborn who introduced me to the life at RSLIS.

Thirdly I want to thank the LIS departments at Humboldt University in Berlin, and Lund University for letting me being a part of their research environments. I have also received important support from the cultural heritage institutions owning the studied resources, especially from Kristine Hoff Meyer at the Danish Agency for Culture, and Jakob Moesgaard and Sigfrid Lundberg at the Royal Library. Thank you for the cooperation.

A special thanks to Franciso Javier Gómez Caballero for the translation of the survey into Spanish.

Last I want to thank my family for their support and patience when I have been occupied nights and weekends. Without them, I would never have been able to accomplish this dissertation.

Lund, March 11, 2014

Abstract

The present thesis investigates the usage of cultural heritage resources on the web. In recent years cultural heritage objects has been digitalized and made available on the web for the general public to use. The thesis addresses to what extent the digitalized material is used, and how findable it is on the web. On the web resources needs to be findable in order to be visited and used. The study is done at the intersection of several research areas in Library and Information Science; *Information Seeking/Human Information Behaviour*, *Interactive Information Retrieval*, and *Webometrics*.

The two thesis research questions focus on different aspects of the study: (1) findability on the web; and (2) the usage and the users. The usage of the cultural heritage is analysed with Savolainen's *Everyday Life Information Seeking (ELIS)* framework. The IS&R framework by Ingwersen and Järvelin is the main theoretical foundation, and a conceptual framework is developed so the examined aspects could be related to each other more clearly. An important distinction in the framework is between object and resource. An object is a single document, file or html page, whereas a resource is a collection of objects, e.g. a cultural heritage web site. Three webometric levels are used to both combine and distinguish the data types: usage, content, and structure. The interaction between the system and its users' information search process was divided into query dependent and query independent aspects. The query dependent aspects contain the information need on the user side and the topic of the content on the system side. The query independent aspects are the structural findability on the system side and the users search skills on the user side. The conceptual framework is summarised in the User-Resource Interaction (URI) model.

The research design is a methodological triangulation, in the form of a mixed methods approach in order to obtain measures and indicators of the resources and the usage from different angels. Four methods are used: site structure analysis; log analysis; web survey; and findability analysis. The research design is both sequential and parallel, the site structure analysis preceded the log analysis and the findability analysis, and the web survey was employed independent of the other methods. Three Danish resources are studied: *Arkiv for Dansk Litteratur (ADL)*, a collection of literary texts written by authors; *Kunst Index Danmark (KID)*, an index of the holdings in the Danish art museums; and *Guaman Poma Inch Chronicle (Poma)*, a digitalized manuscript on the UNESCO list of World cultural heritage. The studied log covers all usage during the period October to December 2010.

The site structure is analysed so the resources can be described as different levels, based on function and content. The results from the site structure analysis are used both in the log analysis and the findability analysis, as well as a way to describe the resources. In the log analysis navigation strategies and navigation patterns are studied. Navigation through a web search engine

is the most common way to reach the resources, but both direct navigation and link navigation are also used in all three resources. Most users arrive in the middle level in ADL and KID, at information on authors and artists. On average cultural heritage objects are viewed in half of the session. In the analysis of the web survey answers two groups of users' are distinguished, the professional user in a work context and users in a hobby or leisure context. School or study as a context is prominent in Guaman Poma, the Inca Chronicle. Generally are pages about the cultural heritage more frequently visited than the digitized cultural heritage objects.

In the findability framework six aspects are identified as central for the findability of an object on the web: attributes of the object, accessibility, internal navigation, internal search, reachability and web prestige. The six aspects are evaluated through seven indicators. All studied objects are findable in the analysis using the findability framework. A findability issue in KID is the use of the secure https protocol instead of http, which leads to the objects in KID having no PageRank value in Google and thereby a lower ranking in comparison to similar objects with a PageRank value. The internal findability is reduced for the objects in top of all three resources, e.g. the first page, due to the focus of the internal search engine on the cultural heritage objects. Several possible adjustment or developments of the findability frameworks is discussed, such as changing the weightning between the aspects measured, alternative scores and automated measuring.

In conclusion, the investigation adds to our knowledge about how resources with digitalized cultural heritage are accessed and used, as well as how findable they are. The thesis provides both theoretical and conceptual contributions to research. The IS&R framework has been adapted to the web, the information search process was split into query dependent and query independent aspects, and a whole findability framework has been developed. Both the empirical findings and the theoretical advancements support the development of better access to web resources.

Abstract in Swedish

Avhandlingen undersöker användningen av kulturarvsresurser på webben. Under senare år har kulturhistoriska objekt digitaliserats och gjorts tillgängliga på webben för allmänheten. I vilken utsträckning det digitaliserade materialet används och hur lätt det är hitta på webben studeras i avhandlingen. Webbresurser måste vara hittbara för att de ska besökas och användas. Studien görs i skärningspunkten mellan flera forskningsområden inom Biblioteks- och informationsvetenskap: *Information Seeking/Human Information Behaviour*, *Interactive Information Retrieval* och *Webometrics*.

Avhandlingens två frågeställningar fokuserar på olika aspekter av projektet: (1) hittbarheten på webben; och (2) användning och användare. Analysen av användningen av kulturarvsresurserna bygger på Savolainens *Everyday Life Information Seeking (ELIS)* ramverk. Ingwersen och Järvelins IS&R-ramverk den viktigaste teoretiska grunden och ett konceptuellt ramverk har utvecklas så att de undersökta aspekterna tydligare kan relateras till varandra. En viktig distinktion är mellan objekt och resurser. Ett objekt är ett enda dokument, fil eller html-sida, medan en resurs är en samling av föremål, t.ex. en webbplats med kulturarv. Tre webometriska nivåer används för att både kombinera och skilja datatyper: användning, innehåll och struktur. Samspelet mellan användare och system i informationssökningsprocessen är uppdelad i frågeberoende och frågeoberoende aspekter. Frågeberoende aspekter är informationsbehovet på användarsidan och ämnesinnehållet på systemsidan. Frågeoberoende aspekter är den strukturella hittbarheten på systemsidan och användarens färdigheter i sökning på användarsidan. Det konceptuella ramverket sammanfattas i *User-Resource Interaction (URI)* modellen.

Forskningsdesignen är en metodologisk triangulering, i form av ett *mixed methods approach* för att få olika bilder av de studerade resurserna och dessas användning. Fyra metoder används: analys av webbplatsens struktur (*site structure analysis*), logganalys, webbenkät och analys av hittbarheten (*findability analysis*). Forskningsdesignen är både sekventiell och parallell, analysen av webbplatsernas struktur föregår logganalysen och hittbarhetsanalysen, och webbenkäten används oberoende av de andra metoderna. Tre danska resurser studeras: *Arkiv för Dansk Litteratur (ADL)*, en samling av litterära texter skrivna av författare, *Kunst Index Danmark (KID)*, ett index av innehaven i de danska konstmuseerna och *Guaman Poma Inch Chronicle (Poma)* ett digitaliserat manuskript som är med på UNESCOs lista över världskulturarv. De studerade loggfilerna omfattar all användning under perioden oktober till december 2010.

Genom analysen av webbplatsernas struktur så kan resurserna beskrivas som olika nivåer, baserat på funktion och innehåll. Resultaten från analysen används både i logganalysen och i hittbarhetsanalysen, liksom ett sätt att beskriva resurserna i sig. I logganalysen studeras navigationsstrategier och navigeringsmönster. Navigation via en webbsökmotor är det vanligaste sättet att nå resurserna, men både direktnavigation och länknavigation används i viss utsträckning

i alla tre resurser. De flesta användare anländer i mittennivån i ADL och KID, där det finns information om författare och konstnärer. Kulturarvsobjekt besöks i genomsnitt i hälften av sessionerna. I analysen av svaren på webbenkäterna har två grupper av användare identifierats, professionella användare i en arbetskontext och användare i ett hobby- eller fritidsammanhang. Kontexten skola eller studier är bara framträdande i Guaman Poma, Inka krönikan. Generellt besöks sidor om det digitaliserade kulturarvet i större grad än de digitaliserade objekten i sig.

Sex aspekter är identifierade som centrala i hittbarhetsanalysen för ett objekts hittbarhet på webben: objektets attribut, tillgänglighet, intern navigering, intern sökning, nåbarhet och webbprestige. De sex aspekterna utvärderas genom sju indikatorer. Resultatet av analysen är att alla studerade objekt är hittbara. Ett problem i KID är användningen av det säkra https-protokollet i stället för http, vilket leder till att objekten i KID inte har något PageRank-värde i Google och därmed en lägre rankning i jämförelse med liknande objekt som har ett PageRank-värde. Den interna hittbarheten är reducerad för objekten i toppen på alla tre kulturarvsresurserna pga. att fokus för de interna sökmotorerna ligger på de digitaliserade kulturarvsobjekten och övrigt ämnesorienterat innehåll. Flera möjliga justeringar eller utecklingsmöjligheter av hittbarhetsramverket diskuteras, t.ex. annan viktning av aspekterna, alternativa poängsättningar och automatiserad mätning.

Sammanfattningsvis så ökar avhandlingen kunskapen om hur webbresurser med digitaliserat kulturarv nås och används, samt hur hittbara resurserna är. Avhandlingen bidrar till forskningen både teoretiskt och konceptuellt. IS&R-ramverket har anpassats till webben, informationssökningsprocessen har delats upp i frågeberoende och frågeoberoende aspekter, och en metod för analys av hittbarhet har utvecklats. Både de empiriska resultaten och de teoretiska framstegen stödjer utvecklingen av webbresursers nåbarhet och hittbarhet.

Table of contents

1 INTRODUCTION 1

1.1 MOTIVATION 1

1.2 OBJECTIVE OF THE THESIS 2

1.3 RESEARCH QUESTIONS 3

1.4 DEFINITIONS 4

1.5 DISCUSSION ON RESEARCH TRADITIONS AND LITERATURE 6

1.6 STRUCTURE OF THE THESIS 7

1.7 EXTERNAL LIMITATIONS 9

2 CONCEPTUAL FRAMEWORK 11

2.1 USER STUDIES 12

2.2 EVERYDAY LIFE INFORMATION SEEKING 17

2.3 OBJECTS AND RESOURCES 24

2.4 INFORMATION SEEKING AND RETRIEVAL 28

2.5 THE INFORMATION SEARCH PROCESS 32

2.5.1 Query-dependent and -independent aspects 34

2.6 THE WEBOMETRIC PERSPECTIVE 35

2.7 THE USER-RESOURCE INTERACTION MODEL 36

2.8 CHAPTER SUMMARY 39

3 CULTURAL HERITAGE OBJECTS AND THEIR FINDABILITY 41

3.1 STRUCTURE AND CONTENT 41

3.1.1 The challenge of the environmental context 41

3.1.2 Linking 42

3.1.3 Content analysis 44

3.2 FINDABILITY 46

3.2.1 Defining Web findability 46

3.2.2 How is something findable on the web? 49

3.2.3 Related web concepts 51

3.3 MEASURING FINDABILITY 52

3.3.1 Object attributes 53

3.3.2 Accessibility 55

3.3.3 Internal navigation 56

3.3.4 Internal search 56

3.3.5 Reachability 57

3.3.6 Web prestige *57*
3.3.7 The total findability indicator *58*
3.4 CULTURAL HERITAGE RESOURCES 59
3.4.1 Definitions of cultural heritage and digital heritage *59*
3.4.2 Digitized cultural heritage on the web *61*
3.4.3 Operationalization *64*
3.4.4 The studied cultural heritage resources *65*
3.5 CHAPTER SUMMARY 67
4 USERS IN ACTION 69
4.1 USER CHARACTERISTICS AND ACTIVITY CONTEXTS 70
4.1.1 User characteristics and information source horizon *71*
4.1.2 Information need, task and intention *72*
4.1.3 Search skills and the web *75*
4.2 INFORMATION SEARCHING AS AN ACTIVITY 78
4.2.1 Modes of searching *79*
4.2.2 Behaviour of web users *80*
4.2.3 Navigational strategies on the web *81*
4.2.4 Paths within the resources *86*
4.3 CHAPTER SUMMARY 88
5 METHODOLOGY 89
5.1 RESEARCH DESIGN 89
5.1.1 The framework and the research design of the project *90*
5.1.2 Measuring the information search process *92*
5.1.3 Triangulation and mixed methods research *93*
5.1.4 Usage and user data *96*
5.2 SITE STRUCTURE ANALYSIS 99
5.2.1 Determining levels of the CH resource *99*
5.3 FINDABILITY ANALYSIS 100
5.3.1 Evaluating findability *100*
5.3.2 The findability measurements *101*
5.3.3 Calculating total, external and internal findability *105*
5.4 LOG ANALYSIS 106
5.4.1 The studied log files *109*
5.4.2 Log pre-processing *112*
5.4.3 Human users versus search engine spiders *113*
5.4.4 Session identification *114*
5.4.5 Measuring path length, visited levels, and arrival level *114*

5.4.6 Measuring the navigation strategy *115*
5.4.7 Analysis of queries in referring search engines *115*
5.4.8 Session paths *116*
5.5 WEB SURVEY 117
5.5.1 Web survey as a method *117*
5.5.2 Survey questions *118*
5.5.3 Deployment of the survey *120*
5.6 MIXING METHODS 120
5.6.1 Data types *120*
5.6.2 Combing different methods *121*
5.6.3 Time overlap in data collection *121*
5.6.4 Potential patterns *122*
5.7 CHAPTER SUMMARY 123
6 THE FINDINGS OF THE SITE STRUCTURE ANALYSIS 125
6.1 THE SITE STRUCTURE OF THE RESOURCES 125
6.2 CHAPTER SUMMARY 128
7 HOW FINDABLE ARE THE RESOURCES? 129
7.1 STUDIED OBJECTS 129
7.2 EVALUATION OF FINDABILITY ASPECTS 132
7.3 EXTERNAL AND INTERNAL FINDABILITY 136
7.4 FINDABILITY AND ITS IMPACT ON USAGE AND NAVIGATION STRATEGIES 139
7.5 REFLECTIONS ON THE FINDABILITY EVALUATION 139
7.6 WEIGHTED FINDABILITY ASPECTS 141
7.7 AUTOMATION OF THE FINDABILITY ANALYSIS 142
7.8 CHAPTER SUMMARY 143
8 THE USE OF THE CULTURAL HERITAGE RESOURCES 145
8.1 HOW USERS ACCESS THE HERITAGE RESOURCES 145
8.1.1 Navigation strategies *145*
8.1.2 Distribution of referring search engines *146*
8.1.3 Queries used in the referring search engines *146*
8.1.4 Referring links grouped per site *149*
8.1.5 Referring Wikipedia pages *150*
8.1.6 Users' countries of origin *151*
8.2 THE SESSIONS 152
8.2.1 Where do the users arrive? *152*
8.2.2 How long do the users stay? *153*
8.2.3 Where do the users go within the resource? *155*

8.2.4 How many leaves directly? *159*
8.3 CHAPTER SUMMARY 161
9 USER CHARACTERISTICS 163
9.1 CHARACTERISTICS OF THE USERS 163
9.2 NAVIGATION STRATEGIES 164
9.3 USERS INTENTIONS AND TASK CONTEXTS 166
9.4 THE USERS AND THEIR USAGE IN THE RESOURCES 167
9.5 COMPARISON OF LOG AND SURVEY DATA 169
9.6 CHAPTER SUMMARY 170
10 DISCUSSION 173
10.1 HOW FINDABLE ARE THE HERITAGE RESOURCES AND THEIR OBJECTS? 174
10.2 HOW DO USERS FIND AND USE THE CULTURAL HERITAGE RESOURCES? 176
10.3 HOW CAN THE DIFFERENT DATASETS BE ANALYSED TOGETHER? 177
10.4 LIMITATIONS OF THE RESEARCH 180
10.5 THE RESULTS IN LIGHT OF ELIS AND IS&R 181
11 CONCLUSIONS 183
11.1 OVERALL CONCLUSIONS 183
11.2 CONTRIBUTIONS TO RESEARCH 183
11.3 IMPLICATIONS FOR PRACTICE 184
11.4 CONCLUDING REMARKS AND DIRECTIONS FOR FUTURE RESEARCH 185
REFERENCES 187
LIST OF ABBREVIATIONS 201
APPENDICES 203

List of figures

Figure 1.1. The sequence of the chapters. 8

Figure 2.1. Wilsons nested model of the information seeking and information searching research areas (Wilson, 1999, p. 263), modified after Ingwersen and Järvelin (2005, p. 198). The legends within the ellipses are from Wilson's original model. 14

Figure 2.2. A revised version of the model from Wilson (1997) with both utilitarian and hedonic outcomes (Laplante, 2008, p. 91). 16

Figure 2.3. Savolainen's model of everyday information practices (Savolainen, 2008, p. 65). 19

Figure 2.4. Information source horizons and zones of source preferences (Savolainen & Kari, 2004, p. 420). 20

Figure 2.5. Information source horizons and information pathways in the context of seeking problem-specific information (Savolainen, 2008, p. 119). 21

Figure 2.6. The Serious Leisure Perspective (SLP) as a map of the leisure universe, showing the three main forms of leisure, their types and subtypes (Hartel, 2009, p. 3265). 23

Figure 2.7. The triangular resource model (a) with the parts of a web resource: navigational objects, informational objects and cultural heritage objects. The circular object model (b) with the objects embedded in the resource, which is available on the web. 25

Figure 2.8. The IS&R framework, a generalized model of participating cognitive actor(s) in context in interactive information seeking, retrieval and behavioural processes, based on (Ingwersen & Järvelin, 2005, p. 261). 29

Figure 2.9. Web IS&R model, a modified version of Bates (2009a); Ingwersen and Järvelin (2005) IS&R model for the present study which takes the complexity of the web into account. The focus in the model is on short-term web interactions, i.e. the model does not include "cognitive transformation and influence" over time, which are included in Figure 2.8. The letters A-D describes different kinds of interactions. 31

Figure 2.10. Bates Berrypicking model of search (Bates, 1989). The model is slightly modified, the arrows between the user on the berrypicking-path and the viewed information objects has been made bidirectional because the object might have impact on the path and the queries. 33

Figure 2.11. The User-Resource Interaction model (URI model) of the interaction between user and the web (including objects and resources) during searching. The elements in the model are marked A-J, which are used as references to specific parts of the model in the text in all chapters. The query-dependent aspects (marked with O) and query-independent aspects (marked with X) are mixed in the information search process as the actions of the user depends on aspects of both types. 37

Figure 3.1. Basic link relations (Björneborn, 2004, p. 16). 42

Figure 3.2. Web node diagram with page level links based on Björneborn (2004, pp. 19-21). 43

Figure 3.3. Illustration of potential kinds of content in the object model (Figure 2.7b). 46

Figure 3.4. Resources and objects from a web perspective. 50

Figure 3.5. The aspects of findability places in the triangular resource model (left) and the object model (right) (Figure 2.7a+b). 59

Figure 4.1. The principle of polyrepresentation of academic documents (Larsen et al., 2006, p. 89). 74

Figure 4.2. Exploratory search activities (White & Roth, 2009, p. 14). 80

Figure 4.3. Illustration of the paths of the different external (1-3) and internal (4-5) navigation strategies in the object-model (Figure 2.7b). 83

Figure 4.4. Navigation strategies with the end point at specific objects in the object model and in the resource model. Note that all the strategies bypass the resource level in the object-model. 84

Figure 4.5. Navigation strategies with the end point at resource level in the object model and in the resource model. 84

Figure 4.6. Navigation strategies with the starting point at the resource level and end point at specific objects in the object model and in the resource model. 85

Figure 4.7. A simplified version of the resource model (Figure 2.7a). 86

Figure 4.8. The six session paths in the simplified two level resource model. The second arrow in B4, N5, O4 and O5 represents one or several pages viewed at the level. 87

Figure 5.1. The URI model (Figure 2.11) is the conceptual framework the research design is built upon. Here with weighted size of the arrows based on their importance in the research design. The letters are the same as in Figure 2.11. 91

Figure 5.2. A multi-level model of contextualized information searching (Hung et al., 2008). The three upper levels are abstract in nature and the lowest fourth level, which is shaded blue, is where the concrete, observable actions of the users takes place. The observable moves (the yellow dots on the fourth level) forms the measurable search process. 93

Figure 5.3. The methods form a methodological triangulation, here presented in relation to each other. The survey is parallel to the other methods, and the site structure precedes the findability analysis and the log analysis in the mixed methods setup. 96

Figure 5.4. The Achecker (http://achecker.ca) test result for http://adl.dk. The number of problems are displayed in the tabs in the middle of the screen dump, e.g. “Known Problems (3)”. . 103

Figure 5.5. An example of the PageRank meter in Google toolbar for Internet Explorer. The page about page in ADL has a PageRank of five out of the maximum ten. 105

Figure 5.6. An example of an access log file entry from (The Apache Software Foundation, 2012). 109

Figure 5.7. An example of line in an ADL log. 111

Figure 5.8. An example of line in a KID log. 111

Figure 5.9. An example of line in a Poma log. 111

Figure 5.10. The 15 session path types in the three level resource model (based on Figure 2.7a). 117

Figure 6.1. The site structure of ADL in the resource-model. 126

Figure 6.2. The site structure of KID in the resource-model. 127

Figure 6.3. The site structure of Poma in the resource-model. 128

Figure 7.1. The number of objects on each level studied in the findability analysis. 129

Figure 7.2. Findability aspects in the resource model and in the object model. The same model as in Figure 3.5, reproduced for for ease of reading. (A. Object attributes. B. Accessibility. C. Internal navigation. D. Internal search. E. Reachability. F. Web prestige.) 133

Figure 7.3. External and internal findability in ADL per level. 137

Figure 7.4. External and internal findability in KID per level. 138

Figure 7.5. External and internal findability in Poma per level. 138

Figure 8.1. The distribution of the navigation strategies in the three resources. 146

Figure 8.2. ADL users' country of origin in the logs. 151

Figure 8.3. KID users' country of origin in the logs. 151

Figure 8.4. Poma users' country of origin in the logs. 152

Figure 8.5. The frequency of session length in KID (n=22666) – a typical distribution. 155

Figure 8.6. The 15 types of paths based arrival level, visited levels and number of pages viewed with the share of each path in ADL. The arrows are only illustrations of the arrival level and the levels visited thereafter in the session. The two values under each path are the share of all sessions (left) and the average session length (right). 156

Figure 8.7. The 15 types of paths based arrival level, visited levels and number of pages viewed with the share of each path in KID. The arrows are just illustrations of the arrival level and the levels visited thereafter in the session. The two values under each path are the share of all sessions (left) and the average session length (right). 157

Figure 8.8. The 6 types paths in a two Poma-level version of Figure 7.4. The paths are based on arrival level, visited levels and number of pages viewed. The arrows are just illustrations of the arrival level and the levels visited thereafter in the session. The two values under each path are the share of all sessions (left) and the average session length (right). 158

Figure 9.1. Staple diagram of answers to the survey question "How did you reach this site?". 165

Figure 9.2. Staple diagram of answers to the survey question "In what context do you visit the web site?" 166

Figure 9.3. Staple diagram of answers to the survey question "Why are you visiting this resource today?". 167

Figure 9.4. Comparison of the distribution of navigation strategies in logs and surveys. 169

Figure 9.5. Comparison of the distribution of country of origin in logs and surveys. 170

Figure 10.1. The indicators placed under the different methods (based on Figure 5.3). 174

Figure 10.2. The three webometric levels in the object model (Figure 2.7b), where the actions at the usage level depends on all the input values into the information search process as illustrated in Figure 2.11, at a specific moment in time. 179

List of tables

Table 3.1. Accessibility and findability in a local e-government study (Kopackova et al., 2010), the division between accessibility and findability is added by me. 52

Table 3.2. Measured aspect of each findability criteria. 52

Table 3.3. Examples of online Danish cultural heritage resource with different dissemination strategies. 63

Table 4.1. Categories and subcategories used in the analysis of queries submitted in the referring web search engine. 75

Table 4.2. Steps in external information navigation strategies (the letters correspond to the arrows in Figure 4.3). 82

Table 4.3. Steps in internal information navigation strategies (the letters correspond to the arrows in Figure 4.3). 82

Table 5.1. Findability measure and points for evaluation of objects attributes in the form of SAPs. 101

Table 5.2. Findability measure and points for evaluation of objects attributes in the form of full text. 101

Table 5.3. Findability measure and points for evaluation of accessibility. 102

Table 5.4. Findability measure and points for evaluation of internal navigation. 104

Table 5.5. Indicator and points for evaluation of internal search. 104

Table 5.6. Findability measure and points for evaluation of reachability. 104

Table 5.7. Indicator and points for evaluation of web prestige. 105

Table 5.8. Findability measurements based on aspect and level. 106

Table 5.9. Explanation of the example in Figure 5.6 from (The Apache Software Foundation, 2012). 110

Table 5.10. The distribution of sessions by human users, search engine spiders and unknown visitors in the log files. 113

Table 7.1. The studied ADL objects. 130

Table 7.2. The studied KID objects. 131

Table 7.3. The studied Poma objects. 132

Table 7.4. Summary of ADL findability evaluations. 134

Table 7.5. Summary of KID findability evaluations. 135

Table 7.6. Summary of Poma findability evaluations. 136

Table 7.7. Findability scores for Poma where the alternative points 0-1-100-1000 are given instead of 0-1-2-3 in Table 7.6. Fulltext is seen as a part of Object attributes and is not awarded separately. 140

Table 7.8. Percentages of total external findability normalised scores depending on different weights in ADL. 141

Table 8.1. Distribution of the two largest referring web search engines (based on Appendix 9). . 146

Table 8.2. The distribution of informational, navigational and transactional queries in the query-sample (based on occurences). 147

Table 8.3. The distribution of the informational subcategories in the query-sample (ADL n=293, KID n=70, and Poma n=60). The number of actual occurances of each subcategory is shown, not their share of the informational queries (see Table 8.4). 148

Table 8.4. Polyrepresentive informational queries (including two or more of the informational subcategories in Table 8.3)..... 149

Table 8.5. Top three referring Wikipedia page in the studied resources (top 10 in Appendix 18). 150

Table 8.6. Number of sessions distributed on Navigation strategy and Arrival level. 153

Table 8.7. Average number of page views per session based on Navigation strategy and Arrival level. 154

Table 8.8. Bounce rate distributed on Navigation strategy and Arrival level in ADL (n=44352). 160

Table 8.9. Bounce rate distributed on Navigation strategy and Arrival level in KID (n=22666). . 160

Table 8.10. Bounce rate distributed on Navigation strategy and Arrival level in Poma (n=30557). 160

Table 9.1. The characteristics of the participants in the survey. 164

1 Introduction

1.1 Motivation[1]

Denmark has increasingly digitized its *Cultural Heritage* (CH) on large scale during the 2000s, which has demanded considerable resources. The collections of digitized materials include substantial amounts of texts, books, pictures and movies. A large part of the collections are made available on the Internet, and the question arises of *how* and *to what extent* these collections are actually used. Another question is if the digitalized cultural heritage corresponds to the public's need for cultural experiences and access to information on the cultural heritage.

Memory organizations are key players in preserving the CH. Archives, libraries, and museums (ALM institutions) are increasingly using the web for publishing CH contents. The CH on the web is accessible through different kinds of sites, from multi-national portal systems, i.e. Europeana, to open collection databases and thematic online exhibitions (Hyvönen, 2012).

The digitized cultural heritage is usually published in a local information system, in a database or in a web publishing system. Sometimes the digitized material buried is far down on the websites of the institutions, making it difficult to find. Sometimes the cultural heritage is launched on theme sites, which are sometimes removed when the project ends. Currently there is a tendency to make cultural heritage accessible through external, non-profit or commercial, online services. The Library of Congress has made images available on Flickr Commons[2], a part of the photo sharing service Flickr where the pictures have free creative common licenses, so the pictures can be reused. The British Museum has had a collaborative project with Wikimedia, the association behind Wikipedia (Wikimedia, 2012). One of the goals of the project has been to disseminate information on the museum's collections to the public. With more and better articles in Wikipedia about the museum's collections, the number of visitors to the museum's website increases. The usage normally increases when there is available information about the digitized cultural heritage in places where the users are (Wikipedia, Google, etc.). The cooperation increases the likelihood that cultural heritage is found by the users (Wyatt, 2010).

Articles on Wikipedia are only one way to increase the digital visibility of cultural heritage sites. Another way is to optimize the site where cultural heritage collections published for the search engines, so-called Search Engine Optimization (SEO). The first step is to allow the material to be indexed by search engines, and that the publishing system or database solution does not prevent

[1] The section is based on a introduction text about the research project (Fransson, 2010).

[2] http://www.flickr.com/photos/library_of_congress/

indexing. In addition to basic requirements of indexability there are many aspects of SEO to end up higher in the hit lists, such as links, keyword frequency in the text and metadata. If the digitized cultural heritage cannot be found in a Google search is the digital visibility is low, as Google is the most important navigation service on the web. Social media may be another important channel for increased digital visibility and increased usage. If the digitized images, sounds, videos and texts are easy for users to share with others, for example with share-buttons linked to Facebook and Twitter, shortcuts in social media leading directly to the cultural heritage can be created. Shared objects which have been relevant to a previous user are a form of social navigation (Dieberger et al., 2000).

The thesis focuses on the public's use of digitized cultural heritage in everyday life. The project studies the digital collections of the memory institutions, operationally defined as the collections that have been saved and then digitized by the archives, libraries and museums in order to limit the investigation. The focus is on the relations between the (search) behaviour of the users, search strategies and the findability of the cultural resources, e.g. how easily the collections are found by the users, and for which purposes the users visits the CH resources.

The lack of an extensive study on the topic motivates the present thesis. The thesis aims to map and analyse the use of the digitized cultural heritage resources in everyday life, as such it seeks to uncover citizens' use of the studied resources. The experience, needs and information behaviour of the users in this respect will involve analyses of search processes and human-system interaction. In recent years large resources have been spent on digitization of cultural heritage and the creation of digital cultural resources, but there has been no analysis of how this digitization serves the citizens. The thesis attempts to remedy this gap and can be described as a part of information behaviour research, which also includes the study of barriers and rejection of the use of digital resources.

1.2 Objective of the thesis

The thesis has two main purposes. The first is to gain an understanding of the usage of the digitized Danish cultural heritage online and its users, primarily in everyday life. This includes the users' information searching behaviour and their intentions and experiences of using digitized heritage resources. The second purpose is to analyse the information searching behaviour in relation to *findability* as the degree of findability of the content may explain some of the search behaviour. The closely related concept of digital visibility is said to be a key driver for traffic to sites in the web (Nicholas et al., 2006c). This part of the project includes the study of how findable the resources are online, e.g. if the resources and their content are indexed by Google.

The thesis will contribute to increased knowledge about how the digitized cultural heritage is found by users, but also increased knowledge of digital information behaviour more generally. Putting the user behaviour in a functional context will potentially be fruitful; the environment can

be a deciding factor in different choice situations. The thesis will study if different types of users can be distinguished according to different information needs and search patterns. Findability as a concept will be developed and applied to the digitalized heritage collections studied, and put in relation to the huge amounts of information on the web. The thesis will also examine if there is a correlation between how much the individual heritage resources and objects are used and their degree of findability.

The findings of the thesis may have implications in many areas such as systems design, information architecture and optimization towards the search engines. Even metadata, copyright issues, and selection for digitization are important aspects, along with guides and audience targeting. The study will increase the knowledge about the users of the digitized cultural heritage, both generally and in relation to the investigated cultural heritage resources.

1.3 Research questions

Information on the web is published in an information system (e.g. content management system, blog or database) and the information system is then available on the web. Information objects have to be found in order to be used. Findability is a prerequisite for usage. Morville defines findability in *Ambient findability* as follows (Morville, 2005, p. 4):

a. The quality of being locatable and navigable.
b. The degree to which a particular object is easy to find or locate.
c. The degree to which a system or environment supports navigation and retrieval.

In Moville's definition findability operates on different levels, both on an object level (b) and on a system or resource level (c). He also discusses the quality of and the degree of findability, which implies that findability can be quantified at some scale. Because of the complexity of web publishing findability is constantly changing and hard to calculate, evaluate or even estimate, but it might be the most important aspect on information on the web in this era of search engine use.

Findability can both be studied within resources, e.g. how easily a object is found inside a resource, and from the web. The two perspectives are important because the users may arrive through different navigation strategies (as discussed below in regard to RQ2) and at different objects in a resource. When arriving in the top of a resource a second, internal search is often needed to for example reach digitized images stored in the resource.

The following research questions address the findability aspects:

RQ1. How findable are the heritage resources and their objects?
- RQ1a. What aspects are important for measuring the findability of a web object?
- RQ1b. How findable are resource and objects from the web?
- RQ1c. How findable are objects within the resource?

The research questions about how easy to find the cultural heritage resources are complemented with a second set of research questions focusing on the users and the usage. Users may reach the cultural heritage resources with different web navigation strategies. On the web there are three basic forms of navigation: direct navigation, navigation through links, and navigation using a search engine (Levene, 2010). The log files contain referring URLs, including search terms from the referring search engine, indicate the information need of the users: informational, navigational or transactional (Broder, 2002). The question is if the users looking for known items or known collections, or if they have general information needs and arrives to the heritage collections in trying to solve them, not looking for cultural heritage in particular, and how the users navigate to the resource.

Research question 2 with sub-research questions addresses the users and the usage of the cultural heritage resources:

RQ2. How do users find and use the cultural heritage resources?
- RQ2a. Which navigation strategies are used by the users to access the resources?
- RQ2b. On what level in the resources do the users arrive?
- RQ2c. How do they navigate within the resources?
- RQ2d. How many objects do the users access in a session?
- RQ2e. Which demographics characterize the users?
- RQ2f. Why do users visit the resources?

The last two sub-research questions addresses aspects of ELIS by analysing the variation of motivation for users to seek digitalised cultural heritage. Due to the two different kinds of data, from users' actions in form of usage data to the content and structure of resources to the findability of the cultural heritage resources, the research is mainly quantitative, but with qualitative elements, and in the overlapping area between several research fields: *Information seeking in everyday life or Human information behaviour* (including *Everyday Life Information Seeking*), *Interactive Information Retrieval,* and *Webometrics.*

1.4 Definitions

Cultural heritage (CH) objects – In the present thesis defined as objects owned and managed by memory institutions (archives, libraries and museums).

Domain knowledge – The user's knowledge about the task, and the topic and context of the task. The domain knowledge is used in combination with the search skills to interact with information resources and objects during the information search process.

Everyday life information seeking (ELIS) – Information seeking and searching in a non-work context. ELIS can range from structured, complex tasks to casual searching for pleasure seeking, and it includes both leisure and non-work tasks

Findability –How easily something is found, for example a specific web page on the web. Findability is seen as an objective and non-domain specific (query-independent) concept measuring structural and contentual aspects of information resources and objects.

Information needs – Are needs for information in a broad sense, which derives from a task and the intention to search for something. Information needs include the whole spectra from casual needs for pastime activities and enjoyment to learning and exploring new topics for complex work-tasks.

Information retrieval (IR) – The actions in the information system during information searching, the term is used for research focusing on system development. A part of the holistic concept of Information seeking & retrieval (IS&R) which comprises all aspects of information seeking, searching and retrieval.

Information search process – The whole process during information search, including both the actions and cognitive activities, the interactions with the information system (digital, physical or human) as well as the feedback from the system, for example in form of new web pages when clicking on links.

Information seeking – The process of fulfilling an information need or complete a task, can be over short or long periods of time. A part of the holistic concept of Information seeking & retrieval (IS&R) which comprises all aspects of information seeking, searching and retrieval.

Interactive information retrieval (IIR) – IIR is an instance of information searching, the interaction with electronic information system. IIR is a development of Information Retrieval (IR) where the user and her behaviour is taken into account when designing and evaluating IR-systems.

Object – An information object is a single item, for example a document or a web page. Resources are collections of objects. Objects might be of different types within a resource, e.g. digitized cultural heritage objects and navigational objects which supports the navigation within the resource.

Query-dependent – Aspects of the information search processes which are connected to the information need, the domain knowledge and the contents of the objects and their subject representation.

Query-independent – Aspects of the information search processes which are connected to general aspects, such as search skills amount of metadata representing the object and the findability of the object or resource.

Resource – A thematic collection of objects, e.g. a web site. The size of a resource may range from a single object (a one-page web site) to thousands of objects (or more). A resource can be described as having different levels where the objects are of specific types, e.g. digitized cultural heritage objects on one level and navigational objects on another level which supports the navigation to the digitized objects.

Search skills – The user's skills in Information seeking and retrieval used during search. The search skills are used in combination with the domain knowledge to interact with information resources and objects during the information search process.

Task – An external problem or quest to be solved through information seeking and searching, the task is the origin of the information need.

Web navigation strategies – Different ways of searching and browsing on the web. The three fundamental web navigation strategies are direct navigation, navigation through links, and search engine navigation.

1.5 Discussion on research traditions and literature

The thesis is based on literature from several research traditions and fields, both pre and post web. User studies and information need research has been a part of *Library and Information science* (LIS) since the 1950s (Saracevic, 2009). With the emergence of the web new possibilities the behaviour of the users and their information needs has emerged, as the all digital actions are possible to record and analyse. *Everyday Life Information Seeking* (ELIS) is a growing research field within LIS, focusing on everyday life and leisure instead of work related information behaviour which is been the dominant focus in user studies in LIS (Savolainen, 2009).

Search strategies is a research area with a long history with roots in several disciplines, e.g. psychology, and has been mixed in LIS. The area has been called "database searching" and is part of a long LIS tradition, with researchers like Raya Fidel (1985) and Marcia Bates (1979). *Human Computer Interaction* was evolving and in the 1990s and had huge impact on research. *Web Information Retrieval* became a central topic and several search services begun as research projects, e.g. Google (Manning et al., 2008). Traditionally *Information Retrieval* (IR) has focused on the computational aspects, which can be criticized for being unrealistic and not taking the user into account. An alternative approach is *Interactive Information Retrieval* (IIR) which brings the user into the IR research. Information seeking and retrieval (IS&R is) an attempt to connect the *Human Information Behaviour* research within LIS with IR (Ingwersen & Järvelin, 2005), both originating in *Information Behaviour* (Wilson, 1999).

Web findability, navigation on the web, and usage of web resources are all three concepts connected to the web. The findability approach is based in the professional fields of *Information Architecture* (IA) and *Search Engine Optimization* (SEO), e.g. Morville (2005) and Walter (2008). These professional fields are pragmatic, based on a mix of research, tacit knowledge and trial-and-error. The academic base of both IA and SEO are LIS and Computer Science, with metadata, Web-IR and web publishing as overlapping research areas.

The thesis is based on several different types of literature. First and foremost it is based on academic research in *Interactive Information Retrieval*, *Information Seeking*, and *Webometrics*;

but also from Computer Science and Internet Research in general. Another important source is the professional literature in the fields of Information Architecture, Search Engine Optimisation, Web Design and Web Analytics. This literature is used to capture best practice and other wisdom of fields not yet been the topic of extensive academic research. These fields are in the broad areas between LIS, Computer Science, Human Computer Interaction, Interaction Design and Web Design. A third kind of literature is from organizations like *World Wide Web Consortium* (W3C) and *International Federation of Library Associations and Institutions* (IFLA) in form of rapports and standards, as well as Danish governmental reports regarding digitalisation of the cultural heritage.

The focus is on cultural heritage resources as a kind of digital resources. The research is not based in cultural heritage research, and the literature concerning the heritage is first and foremost reports from UNESCO and Danish governmental institutions. The research is a study of how the resources are accessed and navigated, as well as their findability. The study will reveal some of the information behaviour of the users of the resources, i.e. the Danish cultural heritage online. The same research approach could be applied to other kinds of resources on the web, it does not depend on the topic of the sites, and for that reason the thesis is mainly based on research from outside the cultural heritage area. The humanistic cultural heritage research often has a philosophical or interpretatative approach (e.g. Lund et al., 2009; *Theorizing digital cultural heritage : a critical discourse*, 2007). The use of the CH or how it is published is not commonly studied.

1.6 Structure of the thesis

The thesis is divided into two main parts, a theoretical and an empirical part. The theoretical part of the thesis is Chapters 2 to 5 and the empirical part consists of Chapters 6 to 9, and it all is wrapped up in Chapter 10 and conclusions are drawn in Chapter 11, as illustrated in Figure 1.1.

Chapter 2 lays the foundation for the whole thesis. The project is based on two different approaches, *Information Seeking and Retrieval* (IS&R) and *Everyday Life Information Seeking* (ELIS), which are moulded together into a framework for combining usage data with content and structural data.

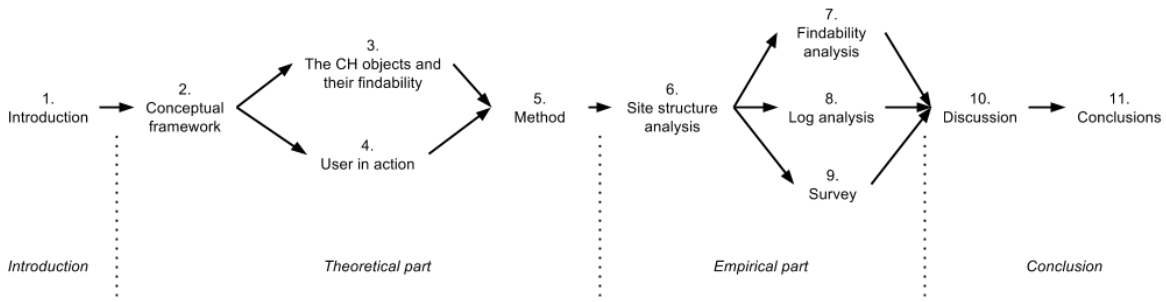


Figure 1.1. The sequence of the chapters.

Chapter 3 focuses on the structural data and on the concept of findability. Web findability is defined and different aspects are discussed. The aspects are related to the models from Chapter 2. The operational findability indicators are discussed. The chapter also focuses on content data and on cultural heritage aspects. The studied cultural heritage resources are presented and discussed in terms of content types. The chapter also has a discussion about cultural heritage resources in general and the limitations of the resources studied.

Chapter 4 focuses on the usage data and metrics derivable from web log files, but it also introduces research that is important for the web survey questions. Navigation strategies to and within web resources are explored in relation to the models discussed in the Chapter 2. Finally, indicators for the usage analysis are presented.

Chapter 5 covers the research design and the applied methods. The research design is described in terms of the models in Chapter 2 and is a mixed methods approach, partly due to the limitations discussed in Section 1.7. Chapter 5 discusses the methods. The main method log analysis is discussed in depth together with surveys on the web. The other two methods, site structure analysis and findability analysis are new methods and was developed for the present study, and are presented as well as critically discussed. In the end of the chapter the first five theoretical chapters are summarized before the empirical chapters in the second part of the thesis.

Chapters 6 to 9 present the empirical investigations. Chapter 6 presents site structure as a method for both describing the resources and as a tool for further analysis as the results are used in chapters 7 and 8. The chapter also presents the results of the site structure analyses of the resources investigated in the present study.

Chapter 7 presents the evaluation of the web findability of the objects in the three CH resources. The evaluation is based on the findability framework developed in Chapter 3 and on the results from the site structure analysis in Chapter 6. The objects which are studied in each resource are listed and a sample of them is shown in the appendices. A summary of the results of the findability analysis of each findability aspect is presented and both external and internal findability are measured by a number of indicators. How the results of the analysis could affect navigation and usage, as well as system design, is discussed as well as how the framework could be improved.

Chapter 8 presents the findings from the log analysis, and discusses and relates them to previous research. Two main aspects are addressed. First how the everyday users access the cultural heritage resources in terms of navigation strategies, queries used in referring web search engines, and referring links. The other main aspects is the sessions, which is studied through session length, visited levels within the resources, and bounce rate.

Chapter 9 covers the results of the web surveys on the three CH web resources. The answers of the respondents indicates thet there are some general usage patterns. A comparison between the survey data and log data is also presented based on the two variables common to both datasets, navigation strategy used and country of origin.

Chapter 10 discusses the empirical findings in relation to the ELIS framework (Chapter 2) and answers the research questions. The limitations of the present research are discussed.

Chapter 11, the last chapter, presents the overall conclusions in regards to the research questions. The contributions to the different research fields and professional practices are discussed, and recommendations for future research based on the results are presented.

1.7 External limitations

The research was partly financed by the Ministry of Culture and the overall topic was defined in the PhD project call. In the call there was several requirements, the project should:

- Study the usage of Danish cultural heritage online;
- Focus on everyday life users (and thereby not specific groups like researchers or hobbyists); and,
- Combine or use several methods, including a HCI approach.

In the project plan in the accepted application eight small projects (cases) were outlined, which has evolved into this multi-method thesis. The research has been kept within the frames drawn by the call together with the project plan, partly because of the tight timetable of three years from start to finish.

The number of Danish cultural heritage resources possible to study is another limitation. Both in terms of how frequently used they are (based on scope and size) and if they were accessible for research.

2 Conceptual framework

> "Conceptual frameworks are best done graphically, rather than in text. Having to get the entire framework on a single page obliges you to specify the bins that hold the discrete phenomena, to map likely relationships, to divide variables that are conceptually or functionally distinct, and to work with all of the information at once."
> (Miles & Huberman, 1994, p. 22)

The use of cultural heritage resources on the web is a complex subject, and therefore a series of perspectives are addressed as a part of the conceptual framework (Sections 2.1-2.5). A conceptual framework is needed to bring a study together, particularly when different methods as well as different types of data are combined. The purpose of the conceptual framework is to relate usage activities with the information environment, and it is the foundation of the research design. The basic notion in the thesis is that users act upon the information available in their environment, both when it comes to books in a physical library (Pors, 1994, 2011) or more generally in a physical information space (Björneborn, 2008, 2011a). In the same manner users interact with information on the web (Pirolli, 2007). Regardless of the nature of the information space, digital or physical, the interaction is about orientation, navigation and movement in combination with the available information or representation of information.

Several approaches are combined into a conceptual framework in the thesis. I have chosen central models as representative for the research area, and which are relevant when studying the usage of resources on the web. An *Information seeking and retrieval* (IS&R) approach is used, which in itself combines *Information seeking* and *Interactive information retrieval* (Ingwersen & Järvelin, 2005, p. VII), to frame the relations studied, the relations between user, information object, information resource and the web (see Section 2.4). *Information seeking* is a part of the *Information behaviour* research tradition and can be either person-oriented research or system-oriented research, but it always focuses on the users (Case, 2007). *Interactive information retrieval* (IIR) on the other hand is a development of the laboratory IR approach which brings the users into the evaluation of IR systems, but the system design is still the main focus (Ingwersen & Järvelin, 2005). The IS&R approach is complemented with a webometric perspective to distinguish different analytic dimensions within the IS&R-relationships in a web context, as both the information seeking and the interactive information retrieval approaches polarises the user and the system without any explicit dimensional distinctions (see Section 2.6). An *Everyday life information seeking* (ELIS) approach based on Savolainen's framework (Savolainen, 2008) is used to put the actions of the cultural heritage users into a larger context (see Section 2.2).

In this chapter the topic of information needs and use is addressed first, then information behaviour and user studies (Section 2.1). ELIS is presented and discussed (2.2). Information objects and information resources are discussed and defined (Section 2.3). The information search

process and IR-interactions are addressed (Sections 2.4 and 2.5) and leading to a three dimensional conceptual framework. In the chapter I will develop a model of the interactions between users and web objects during search with three levels of interaction: structure, content and usage (Section 2.6). The model covering the User-Resource Interaction (URI, see Section 2.7) is used in the thesis both as a theoretical and conceptual cohesive force, and as foundation of the research design.

2.1 User studies

> "*Information need* refers to a cognitive or even a social state and *information use* to a process." (Saracevic, 2009, p. 2577)

In the research tradition of Library and Information Science the thesis is a "user study", as it studies users and their actions. In one sense the whole conceptual framework is my positioning within the field of user studies and a clarification of my theoretical standpoint.

The study of the behaviour of the users in the context of information needs and information systems has a long history in Library and Information Science. Different concepts and terms have been used over time for the intentions and actions of users. The concept of *information need* was, according to Saracevic, used before the 1980s as a primitive concept on two levels: (1) on the individual level the information need was a cognitive state underlying the information requests; and (2) on the social level it was imaged to correspond to the information requirements of whole groups, e.g. chemists, which could be satisfied with specific information sources. After 1980 the concept was increasingly criticized and at the end of the century the concept of information need was largely abandoned. The research focused on information seeking and other aspects of information behaviour, instead of information needs (Saracevic, 2009). Case discusses information needs and has divided the recent views of the concepts into three categories: seeking answers; reducing uncertainty; and making sense (Case, 2007, pp. 72-76). The categories not only highlights different research approaches taken, but they also suggest that in the extension of the information need there may be different types of seeking behaviour, which will be addressed below. Case concludes that the concept of information needs is "awkward […] particularly in that it is not easily observable" (2007, p. 81).

Information use is as a concept more precise than information need and is possible to study through observation. The information use studies are pragmatic, retrospective, and descriptive by nature (Saracevic, 2009). Saracevic describes the study of information use as:

> "In information science, information use refers to a process in which information, information objects, or information channels are drawn on by information users for whatever informational purpose. The process is goal-directed. Questions are asked: *Who are the users of a given information system? What information objects do they use? What*

> *information channels are used to gather information?* Or in other words: *Who uses what? How? For what purpose?"* (Saracevic, 2009, p. 2578)

The study of information use was in the beginning, in the 1950s, focused on users in the fields of science and technology. Over the years the focus was expanded to include other groups of users, and in the 1990s even the use in everyday life was included (Saracevic, 2009). Recent examples of use studies are the use of a social tagging system (Heckner et al., 2009) or the motivations behind the use of Facebook (Joinson, 2008). At the same time the study of information use might be replaced by the emerging field of *Human Information Interaction* (HII) as the information objects becomes more ambiguous and changing in real time during the interactions (Marchionini, 2008). Information behaviour is another term frequently used and has a slightly broader meaning as it encompasses both search and use behaviour:

> ""Information behavior" is the currently preferred term used to describe the many ways in which human beings interact with information, in particular, the ways in which people seek and utilize information." (Bates, 2009b, p. 2381)

But the concept is not unproblematic, Savolainen argues for the concept of *information practices* instead:

> "The concepts of information behavior and information practice both seem to refer to the ways in which people "deal with information." The major difference is that within the discourse on information behavior, the "dealing with information" is primarily seen to be triggered by needs and motives, while the discourse on information practice accentuates the continuity and habitualization of activities affected and shaped by social and cultural factors." (Savolainen, 2007, p. 126)

The differentiation between the concepts stresses the different epistemological approaches and creates two research areas instead of one. It could be argued that during system interaction internalized social and cultural factors are a subset of all the cognitive and affective factors, except in situations where there are an obvious external pressure on the user, i.e. during collaborative searching. Neither information behaviour nor information practices address the information space as a context of nor an influencer on the information behaviour (discussed in Chapter 4). *Information seeking* has been a major focus of research in both *Human information behaviour* (HIB) and *Information practices* (Case, 2007).

> "Information seeking, as is the case with most human information behavior, is highly dependent on context." (Saracevic, 2009, p. 2578)

Information seeking and *information searching* are often used interchangeably, which reduces the usefulness of the concepts. However, Wilson has proposed an often cited and useful model of the three concepts information behaviour, information seeking and information searching (Figure 2.1) on which I base my understanding of the phenomena.

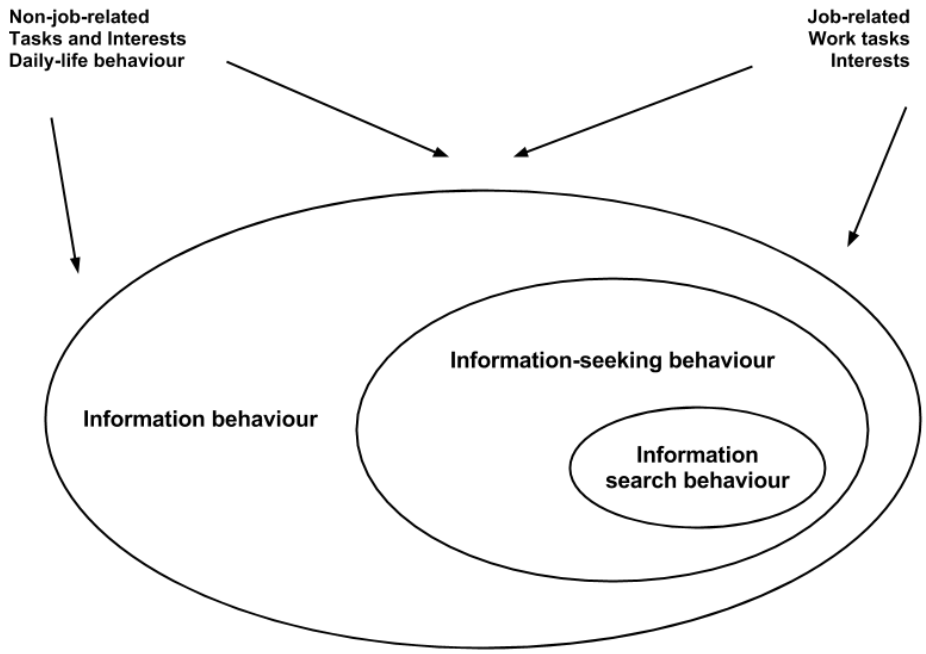


Figure 2.1. Wilsons nested model of the information seeking and information searching research areas (Wilson, 1999, p. 263), modified after Ingwersen and Järvelin (2005, p. 198). The legends within the ellipses are from Wilson's original model.

Information search and navigation are in the thesis viewed as Wilson's *information search behaviour,* i.e. actions and interactions during a session, and information seeking is used as a broader concept which might coincide with searching or cover several search episodes (Wilson, 1999). Saracevic has a similar definition of the two concepts:

> "Information searching is a subset of information seeking, and in the context of information science, it refers to processes used for interrogating different information systems and channels in order to retrieve information. It is the most empirical and pragmatic part of information seeking studies. Originally, search studies concentrated on observation and modelling of processes in the interrogation of IR systems. With the advent of digital environments, the focus shifted toward Web searching by Web users. New observational and experimental methods emerged, becoming a part of exploding Web research. Such search studies have a strong pragmatic orientation in that many are oriented toward improving search engines and interfaces, and enhancing human–computer interactions." (Saracevic, 2009, p. 2579)

There are numerous of models of information seeking and information searching within LIS. Several authors compare models based on different perspectives, e.g. type of model (Fidel, 2012), search models (Xie, 2009), for web search (Knight & Spink, 2008) and exploratory search (White & Roth, 2009). Fidel (2012) divides the models into three different types:

> "Models in ISB [information seeking behavior] can be divided according to the dimension of reality they represent: *action models* represent activities during information seeking and, at times, even before and after; *element models* represent elements that shape information seeking (or, to translate into positivist language: models that represent the variables affecting information seeking). Other models—mixed models—include both; some side by side, others in an integrated fashion." (Fidel, 2012, p. 64)

Fidel discusses some models of each type, examples of action models are Kuhlthau's model of the Information search process (ISP) (e.g. 1991) and Ellis (1993). Belkin (e.g. 1982) and Ingwersen (e.g. 1996) are two of the element models discussed by Fidel, and among the mixed models are Savolainen (1995) and Wilson's second model (e.g. Wilson, 1997). The latter models often represent activities in the seeking process, and then link the activities to the elements that influence or shape them (Fidel, 2012). Laplante studied music information seeking behaviour and used Wilsons second model, but she revised it in several ways (Laplante, 2008). The types of information seeking behaviour originally in the model were replaced with Bates four modes of searching (being aware, monitoring, browsing, and searching) (Bates, 2002) and most important, the possible outcomes are two. To the original outcome in the model, Utilitarian (Information processing and use), Hedonic (Pleasure) has been added to capture ELIS aspects like entertainment (Figure 2.2). Laplante's revised Wilson-model is thus more suited to address information behaviour in both work and non-work contexts.

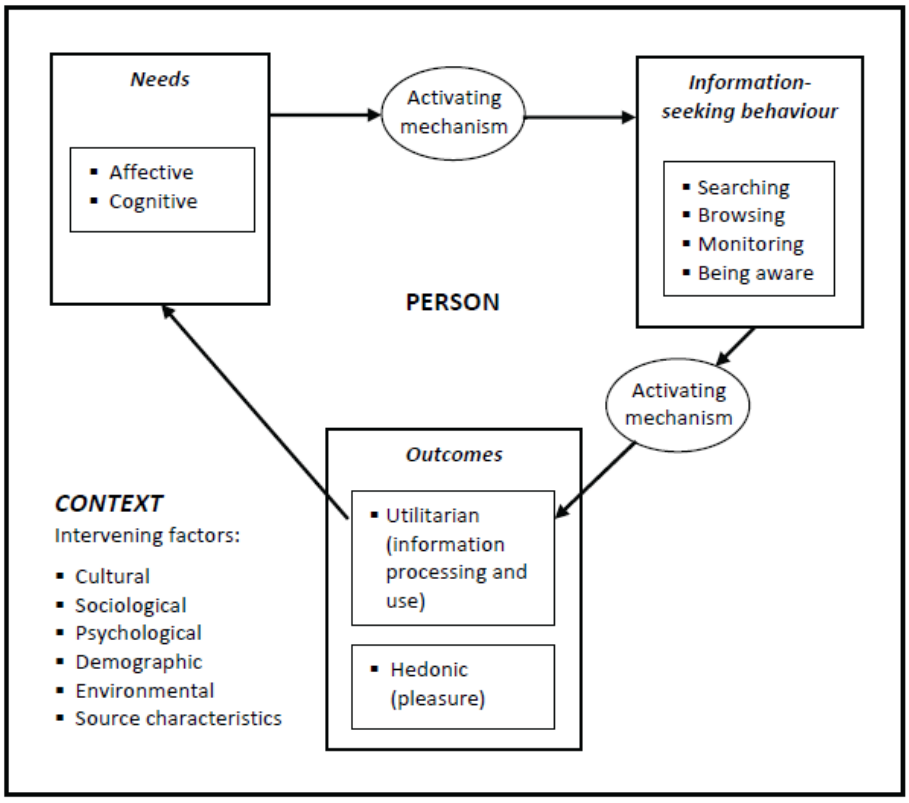


Figure 2.2. A revised version of the model from Wilson (1997) with both utilitarian and hedonic outcomes (Laplante, 2008, p. 91).

Information searching can also be of a different kind in terms of the goal of the search, which is not taken into account in Laplante's model. Different needs or intentions generate different types of searches. A need to confirm a fact is different from the need to learn a new subject, and the two needs are fulfilled in different ways. The traditional information retrieval systems are based on a look-up model, where the query of the user is matched with the representations of information objects. The look-up systems are best suited for fact finding or question answering (White & Roth, 2009). Marchionini has formulated the problem with information searching through query formulation and introduced the concept of *information interaction*:

> "A person with an information problem is best able to meet that need through action, perception, and reflection rather than through query statements alone. Thus, the notion of information interaction rather than information retrieval to better reflect the active roles of people and the dynamic nature of information objects in the electronic environment." (Marchionini, 2006b, "Conclusion")

In the thesis *information search process* is used instead of *information interaction* because it stresses both search and is a goal-oriented process. Information interaction was introduced as a reaction to the concept of information retrieval, not information searching or information seeking. Information interaction is a part of the evolving field Human Information Interaction (HII), which changes the focus of Human Computer Interaction (HCI) from interaction with a computer to

interaction with information (Marchionini, 2008). In the thesis information interaction is seen as a part of the information search process.

A solution to the problem Marchionini addresses in the quote above is to develop systems that support exploratory searching. *Exploratory searching* is defined as:

> "Exploratory search can be used to describe an information-seeking problem context that is open-ended, persistent, and multi-faceted; and to describe information-seeking processes that are opportunistic, iterative, and multi-tactical. In the first sense, exploratory search is commonly used in scientific discovery, learning, and decision-making contexts. In the second sense, exploratory tactics are used in all manner of information seeking and reflect seeker preferences and experience as much as the goal." (White et al., 2008, p. 433).

Exploratory search is addressed in Section 4.2.1. Another issue in the information behaviour research area (*User studies*) is addressed by Saracevic:

> "By 2008 there are still two worlds of user studies: one more pragmatic, but now with the goal of providing the basis for designing more effective and usable contemporary IR and Web systems, including search engines, and the other more academic, still with the goal of expanding understanding and providing more plausible theories and models. The two worlds do not interact well." (Saracevic, 2009, p. 2578)

Fidel has made a similar remark, but talks about research in Human Information Behaviour (HIB) and Interactive Information Retrieval (IIR) instead of academic and pragmatic user studies. The research results in HIB are mainly descriptive. In IIR the models are normative, and it is hard to combine the descriptive results from HIB into IIR improvements (Fidel, 2012). These two worlds, in Saracevic's words, have proved to be challenging to combine in one conceptual framework, and it is one of the reasons that I will develop a new model of the interaction between user and web resource (Figure 2.11). The model is an attempt to bridge the model-gap between HIB and IIR, where the "descriptive" behaviour of the users and the "normative" information system are related to each other.

This section on user studies has clarified my use of some central concepts, and has pointed out several issues within the fields of user studies. The long term focus on utilitarian outcomes, the new challenges addressed with exploratory searching, and the gap between the HIB models and the IIR models are aspects of user studies that will be addressed in the thesis.

2.2 Everyday life information seeking

It is necessary to include an everyday life perspective because part of the usage of the CH resources is in non-work contexts. It is also an outspoken goal with the digitization of the Danish CH that the collections of CH objects should be available to the citizens. *Everyday Life*

Information Seeking (ELIS) is a framework for studying information behaviour and information practices in everyday life. It has been developed as an alternative to the great focus on work as a context and work tasks as the incentives for information seeking (Savolainen, 2009). Savolainen focuses on the everyday life tasks and the routines, and has studied both unemployed people and environmentalists (Savolainen, 2008).

> "Terminological problems originating from the false dichotomy of work-related and "nonwork" information seeking may be avoided by taking the concept of ELIS as starting point. The key word is *everyday life*, which refers to a set of attributes characterizing relatively stable and recurrent qualities of both work and free time activities. The most central attributes of everyday life are familiar, ordinary, and routine, and they qualify the structural conditions of action (e.g., the recurrent "rhythms" of work and leisure hours). The above characteristics of familiar, ordinary, and routine become real only in the process in which they are reproduced, day after day." (Savolainen, 2009, p. 1781)

Savolainen has moved from a Foucault-inspired view of ELIS to a phenomenological ELIS with focus on practices. As the ELIS framework has evolved the early ELIS model (Savolainen, 1995) has been replaced by a series of models and illustration (Savolainen, 2008). How information practice relates to information behaviour has been debated. Savolainen sees the two concepts as parallel umbrella concepts (Savolainen, 2007), while Wilson views behaviour as a generic concept and practice as an element of behaviour (Wilson, 2008) when they debated after Wilson's book review of Savolainen (2008). Two comments in the subsequent debate were "Practices are made up of behaviours" and "the two different concepts doesn't differ in any significant way" ("The behaviour/practice debate: a discussion prompted by Tom Wilson's review of Reijo Savolainen's Everyday information practices: a social phenomenological perspective.," 2009). In the present thesis information behaviour and information practice are seen as similar concepts, and the distinction between them is not central as the main focus is on system interaction. ELIS is used as an analytic framework for interpretation of the findings in an everyday context, and the IS&R framework serves as a conceptual foundation (see also Section 2.1).

The concept of information practice described by Savolainen also consists of information use and information sharing, besides information seeking (Figure 2.3). Due to the focus on (social) information practices it may be argued that *Personal Information Management* (PIM) (e.g. Jones, 2008, 2009) aspects are lacking in the model. Storing and organizing information are important PIM activities besides finding, using and sharing.

The ELIS activities are carried out within a broader context of everyday practices and projects. This broad and general context of the information seeking, using and sharing activities is the totality of individual experiences as well as the transindividual life-world, which is based on social, cultural and economic factors. There are also contextual factors affecting the ELIS activities, like time constraints and the relative importance of the everyday project at hand (Savolainen, 2008). Both contexts are integrated into Savolainen (2008) model of everyday information practices in Figure 2.3.

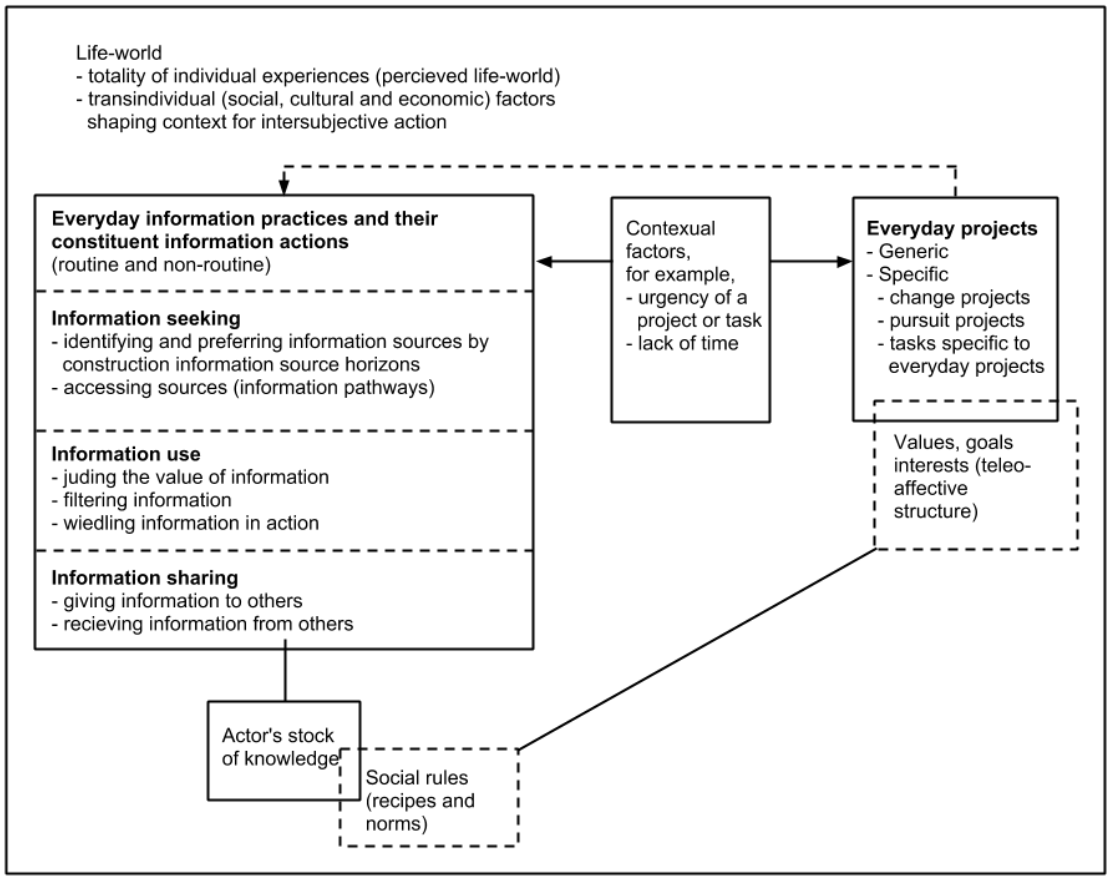


Figure 2.3. Savolainen's model of everyday information practices (Savolainen, 2008, p. 65).

The basis of the information practice is the actor's stock of knowledge (placed under the Everyday information practices in Figure 2.3), because "the ways in which information is sought, used, and shared draws on this resource" (Savolainen, 2008, p. 66). As the actor gains new experiences her knowledge is broadened and deepened, and her practices becomes more refined. The refined information practices are tools in everyday life to serve the furtherance of the different everyday projects, not a goal in itself (Savolainen, 2008). The knowledge the information practices are based on is discussed in Chapter 4, especially the search skills.

Savolainen does not distinguish searching from seeking. He focuses on the sources the user visits, the paths between them (Figure 2.5), and how the sources are placed within the information source horizon of the user (Figure 2.4) (Savolainen, 2008). Two important concepts in ELIS are *information source horizons* and *information pathways*, and they are defined by Savolainen as:

> "The ways in which people identify and access information sources are oriented by their *information source horizons*, that is, the ways in which information sources are perceived to be available in different situations such as monitoring daily events or solving specific problems. The preference (or avoidance) of information sources is based on such perceptions. Further, information seeking is affected by the nature *of information*

> *pathways*, that is, the sequence in which the (preferred) sources are accessed." (Savolainen, 2008, p. 50)

There are two types of information source horizons. The relative stable horizons which indicate how people tend to value information sources over time and across situations, and the dynamic horizons which are problem- or situation specific (Savolainen, 2008). How an information source horizon is constructed is described by Savolainen:

> "When constructing an information source horizon, the actor judges the relevance of information sources available in the information environment and selects a set of sources, say to clarify a problematic issue that is important to the everyday project. Thus, due to the selective approach to information sources, the horizon covers only a part of the actual information environment. Finally and most importantly, the selected information sources are positioned preferentially within the horizon so that the most relevant ones will be placed closest to the actor and the least relevant farther on." (Savolainen, 2008, p. 61)

In the quote Savolainen explains that the sources are positioned by perceived relevance, but empirical research has found that familiarity and accessibility are other key factors for the position of a source within an information source horizon (Savolainen & Kari, 2004).

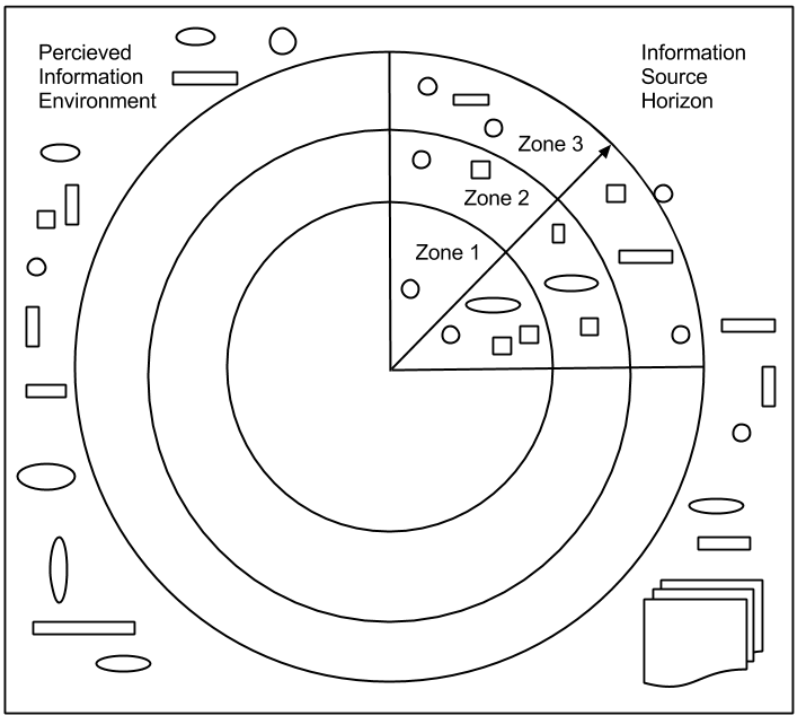


Figure 2.4. Information source horizons and zones of source preferences (Savolainen & Kari, 2004, p. 420).

The number of zones in an information source horizon is arbitrary and is in Figure 2.4 illustrated types with three zones (Savolainen & Kari, 2004):

Zone 1 = Most strongly preferred information sources
Zone 2 = Information sources of secondary importance
Zone 3 = Peripheral information resources

The zones of source preference are important because during ELIS people tend to use only a few source types and a major factor in choosing source is its perceived accessibility, especially in Zone 1 the "principle of least effort" has been found important (Savolainen & Kari, 2004).

The order in which the actor uses the sources forms an *information pathway* (Johnson et al., 2006). The dynamic information pathways complements the concept of information source horizons which suggests "a fairly static approach in that it stands for the constellation of source preferences" (Savolainen, 2008, p. 63). As shown in Figure 2.5 the information pathway consists of the visited sources in the different zones in ascending order, i.e. the sources in Zone 1 are visited before the sources in Zone 2.

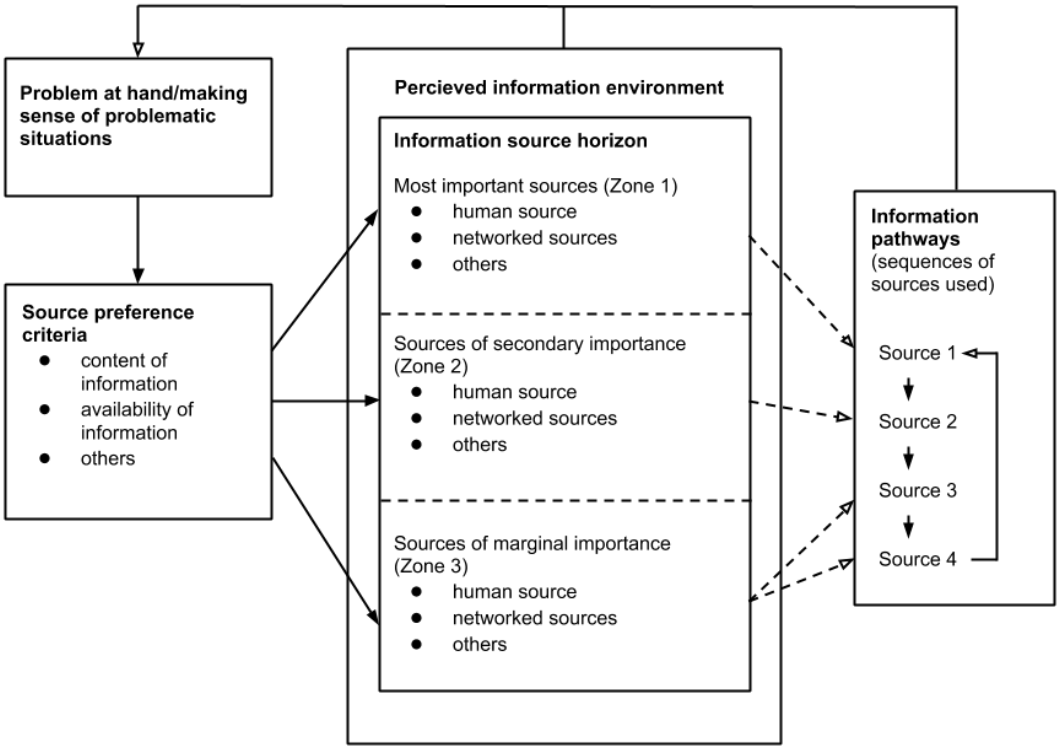


Figure 2.5. Information source horizons and information pathways in the context of seeking problem-specific information (Savolainen, 2008, p. 119).

The model in Figure 2.5 deals with the context of "seeking problem specific information" and it is complemented in Savolainen (2008) with a model for the context of "seeking orienting information" for passive information behaviour. In the ELIS framework the use of cultural heritage resources seen as active searching, and thereby the second model is ignored in the thesis. All the models and concepts are integrated in Savolainen's view:

> "Information source horizons and information pathways are constitutive of the practice of information seeking, since these constructs orient the ways in which information sources are set in preferential order and accessed with regard to their potential usefulness […]. The practice of information use is oriented by the judgement of the value of information. From this perspective, information is filtered to identify the most relevant content that may be wielded in the furtherance of everyday projects […]. Finally, the practice of information sharing is constituted by actions that stand for giving information to others, and receiving information given by others." (Savolainen, 2008, p. 65)

As stated by Savolainen in the quote above are the concepts in Figure 2.3, Figure 2.4, and Figure 2.5 in interplay in the information search process (as illustrated in Figure 2.11). When it comes to cultural heritage objects their findability I observe a part of the horizons and the pathways in the log files of the CH resources studied, and another part in the survey answers by users of CH resources. The findability of the objects may influence the horizons, the pathways taken and the usage of the objects, but the content-type of the objects must also match the needs of the users for the CH object to be interest of the users.

One aspect of ELIS is serious leisure as Stebbins (2007, 2009) calls his framework. Within serious leisure there are various degrees of seriousness, from passtime activities to limited projects like planning a wedding, to long term hobbies demanding considerable resources. The three overall categories are, as shown in Figure 2.6: Casual leisure, Project-based leisure, and Serious leisure. One example of the latter category was studied by Hartel (2010). She examined how amateur gourmet chefs organised their collections of cookbooks and recipes. The distinction in essence between the three types is captured in the following quote:

> "Whereas casual leisure supplies pleasure, and project-based leisure delivers a temporary reward, serious leisure generates deep and enduring sensations of fulfilment." (Hartel, 2009, p. 3267)

Serious leisure is divided into three main types: amateurism, volunteering and hobbies. Amateur has professional counterparts and operates within for example sports and arts. Volunteering is informal or formal help without payment, which may be relevant in connection with cultural institutions like local museums. Hobbies are systematic and enduring pursuit of an activity that leads to acquisition of knowledge, skills, and experiences, and are therefore information-rich (Hartel, 2009).

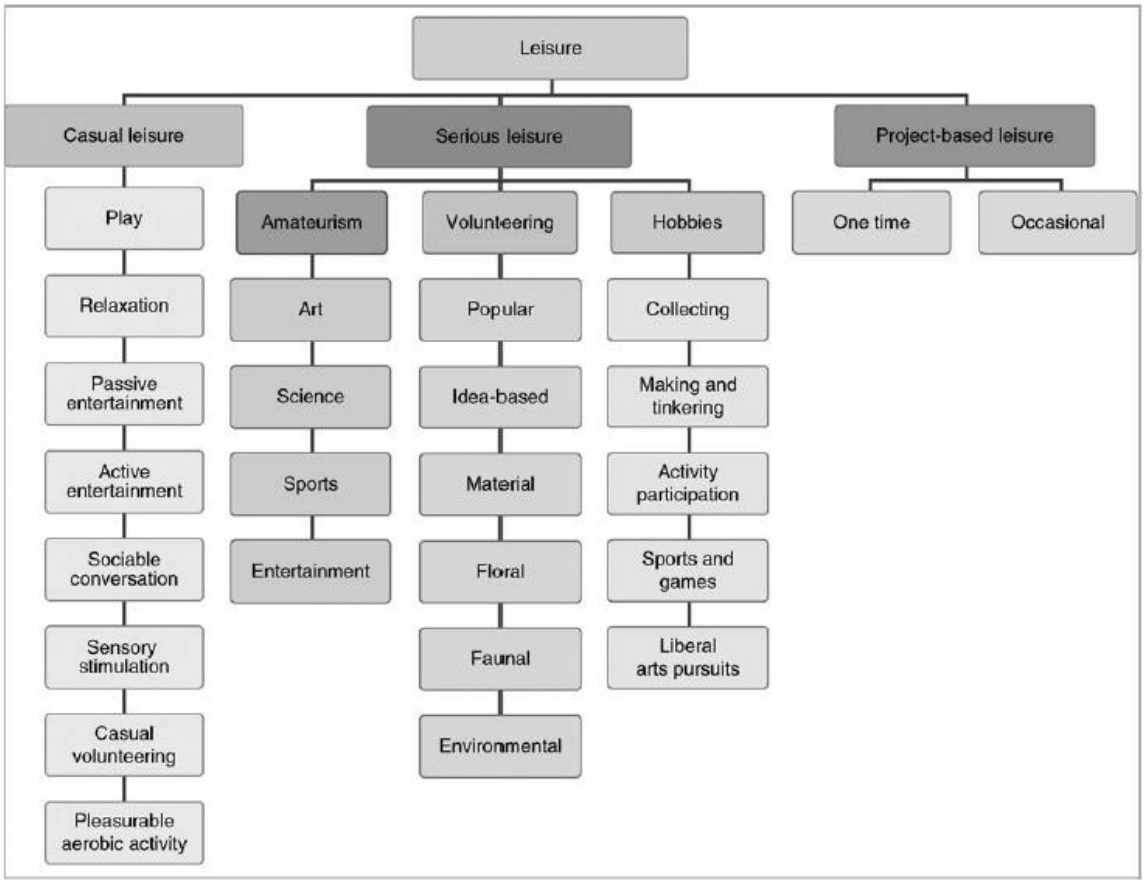


Figure 2.6. The Serious Leisure Perspective (SLP) as a map of the leisure universe, showing the three main forms of leisure, their types and subtypes (Hartel, 2009, p. 3265).

In the context of using cultural heritage resources on the web several of the types and subtypes of leisure in Figure 2.6 are relevant. Surfing the Internet is seen as "passive entertainment" as thirty percentage of the users go online for non-purposive Web surfing (Fallows, 2006; Hartel, 2009), and some users visits cultural heritage web sites non-purposely. The cultural heritage resource could be used as a part of project-based leisure, for example for vacation planning or preparation before a museum visit. Several subtypes of "serious leisure" like art amateurism, volunteering in cultural institutions or cultural historical societies, or liberal arts pursuit might be relevant in regards to cultural heritage usage.

Advanced project-based leisure projects like planning a wedding are far from Savolainen's ELIS which is focused on routines because few amateurs will incorporate the information seeking to be part of their daily routines in the long run (e.g. after the wedding). Similarly many work tasks in a work context can be described as routine rather than unique and complex. For example physicians often control side effects for medicine, an easy task, while larger and more complex search tasks are largely avoided (Hersh, 2009). Savolainen takes this duality into account, in terms of the false dichotomy between work and non-work related information seeking in the quote beginning of the section, where he takes everyday life as a starting point, not non-work (Savolainen, 2009, p. 1781). Work- and everyday life perspectives on information seeking are

complements rather than mutually exclusive, at least when the interaction between user and information/system is studied.

The central contribution of ELIS to the conceptual framework is the focus of the individual user and her information searching, especially the concepts information source horizon and information pathways. The ELIS concepts makes it possible to analyse and discuss the usage of the CH resources in the larger contex of every day life activities, but also in work contexts as the ELIS concepts can be applied in other contexts as well. The Serious Leisure concepts are important for the understanding the reasons for the use of the CH resources in every day life, as they complement works tasks and study assignments as motive for searching. Serious Leisure is an extention of the hedonic outcomes discussed in Section 2.1. This section on ELIS is central for understanding the use of the CH resources (see the discussion of the findings in Chapter 10).

2.3 Objects and resources

All information on the web is published in some sort of information resource, which is made public on a web server. How findable the information is depends on many things, but the user always has to find her way on the web to the resource containing the information objects. Objects are the single pictures, videos, texts, etc. containing the cultural heritage. On the resource level there are navigational functions like internal search and categories together with general information. The resource level is the folding around the collection of objects, and is a layer that has to be penetrated to reach the objects. The distinction is important when looking on search strategies and navigational paths on the web. With the distinction between different types of objects within a resource it is possible to study the users' interactivity with the different types of objects. The differentiation creates analytic categories, with which it is possible to study how visited for example digitalized CH objects are in relation to the types of objects within a resource without studying each individual object.

The concepts of resource, object, and types of objects are further developed into two models in the next section to clarify and visualise their relationships. Within the cultural heritage resources there may be different kinds of objects. The first distinction is made between objects that are the digitized cultural heritage (CH objects), i.e. the objects which are the resource's goal to mediate to the public, and the non-cultural heritage-objects. The non-CH objects are pages with general information about the resource, internal search, etc. There could also be pages with additional information about the CH objects and e.g. their creators. Three types of objects are defined as follows and are positioned in Figure 2.7:

1. *Cultural Heritage objects* are objects which are the goal of the resource to make public;
2. *Informational objects* are objects with additional information, e.g. about the artists or authors of the CH objects, and;

3. *Navigational objects* are of a general character, e.g. the top page, a general "about" page, topical categories and internal search engine, not associated directly to specific CH-objects or their creators.

A CH resource is often made up of the cultural heritage objects, of navigational pages, like homepage and pages with browsable categories and internal search, and of informational objects concerning the CH objects. A resource is a collection of objects within a technical system, which includes thematically and structural aspects. Every site or resource has an internal structure, the functions and contents are generally organized in a hierarchal manner. General information about the site is often close to the top page; whereas more specific information is placed further down in the structure. Based on a need for structuring the objects within the CH resources in order to be able to analyse both the usage in form of navigation within the resource and the findability of different object types I have created two different models of the relations between objects, resource and the web (Figure 2.7).

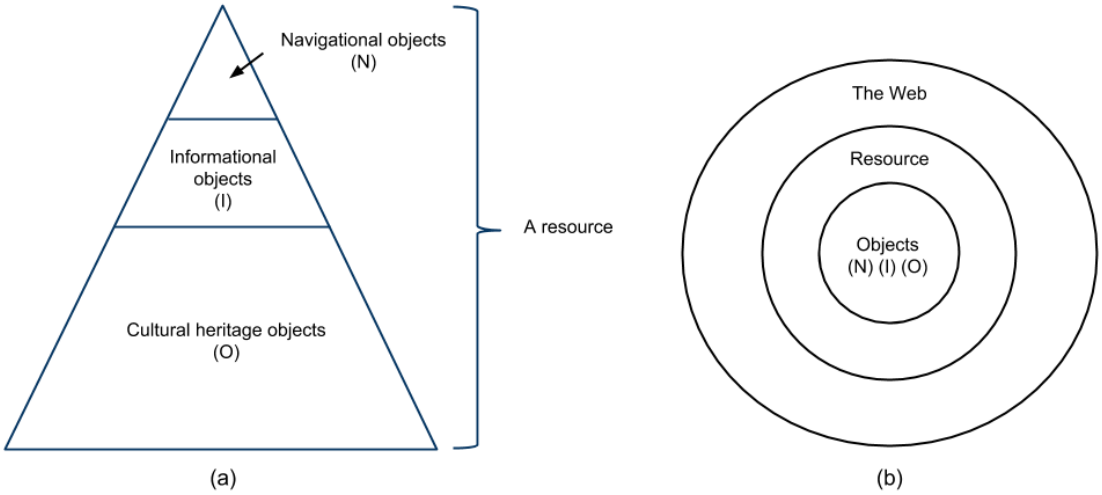


Figure 2.7. The triangular resource model (a) with the parts of a web resource: navigational objects, informational objects and cultural heritage objects. The circular object model (b) with the objects embedded in the resource, which is available on the web.

In the triangular resource model (Figure 2.7a) a resource is viewed as a triangle with different levels. The different types of pages/objects in the CH resources can be categorized as belonging to one of the three levels. The navigational pages (N) consist e.g. of the top page, about the site pages, internal search engine and link categories. The informational pages (I) are pages about the cultural heritage objects. The CH objects (O) are the actual texts and artworks, the large base of digitized objects. The levels above are just an example of levels based on a typical cultural heritage resource. The number of levels is based on the content and structure of the resource and on the research goal. I have created both the categorisation of objects and resources, and the models in Figure 2.7 based on observation and analysis of the CH web sites. A similar approach was used by Montgomery and colleagues when they studied paths in an online bookstore, and they classified the pages into seven categories (Montgomery et al., 2004). Implicitly the Ciber

research group[3] has used a similar division between different levels, for example in their analysis of Europeana, the European cultural heritage portal (Clark, 2011).

A different view is provided by the circular object model (Figure 2.7b). This allows a focus on the objects, which are placed in the center of the model and they are accessible through the resource, which is available on the web. The object in the center might be of any kind, navigational, informational or cultural hertige.

The term "resource" is not unproblematic, neither is the term "object". "Document" is often used in IR and LIS (Buckland, 1991, 1997), but what is a document when it comes, e.g. to a digitized book, the book itself or all the image files of each page in the book? In *Functional Requirements for Bibliographic Records*, *FRBR* (IFLA, 2008), there are four main concepts concerning artistic products: *Work* is "a distinct intellectual or artistic creation." (IFLA, 2008, p. 17); *Expression* is "the intellectual or artistic realization of a *work* in the form of alpha-numeric, musical, or choreographic notation, sound, image, object, movement, etc., or any combination of such forms." (IFLA, 2008, p. 19); *Manifestation* is "the physical embodiment of an *expression* of a *work*." (IFLA, 2008, p. 13), and; *Item* is "a single exemplar of a *manifestation*." (IFLA, 2008, p. 24). If the described document is a digitized version of a unique manuscript, the physical manuscript is, in the terms of FRBR, a work, an expression, a manifestation and an item, all at the same time. The four concepts are coinciding in the case of the physical document. When the manuscript is digitalized the digital version constitutes a new manifestation of the same expression as the physical version. If the digital version is in the form of a pdf-file the single file constitutes an item, according to the definition above. But if the manuscript is in the form of 200 picture files then the collection of files does not constitutes an item, they are not "a single exemplar of a manifestation" (IFLA, 2008, p. 24). The FRBR framework cannot handle this type of digitalization of physical objects. But these different levels of granularity are crucial on the web.

Kallinikos, Aaltonen and Marton has developed a theory of digital objects. Digital objects differ from physical objects in four dimensions, they are: (1) Editable, a least in principle they can be modified; (2) Interactive, "in the sense of offering alternative pathways [...] or explore the arrangements of information items underlying it and the services it mediates." (Kallinikos et al., 2010, "Digital objects: Definitions and attributes"); (3) Open, possible to access and modify by other digital objects, e.g. a web browser or a picture editing software; and (4) Distributed, in the sense that they seldom are contained within a single source or institution (Kallinikos et al., 2010). The shortcomings of FRBR in the case of digital objects are summarized as follows:

> "Most crucially, they [digital objects] are assembled into units by operations that are technologically driven and frequently far beyond the desktop by which users access or manipulate them. Accordingly, their evasive identity raises problems of authentication

[3] http://ciber-research.eu/

> and preservation and impinges upon the inherited functions and practices of memory institutions like libraries and archives." (Kallinikos et al., 2010, "Preamble")

The World Wide Web Consortium (W3C) and the founder of the web Tim Berners-Lee use the term "resource", e.g. in Uniform Resource Identifier (URI) and Uniform Resource Locator (URL). Two definitions of resource are:

> "A "resource" is a conceptual entity (a little like a Platonic ideal). When represented electronically, a resource may be of the kind which corresponds to only one possible bit stream representation. An example is the text version of an Internet RFC. That never changes. It will always have the same checksum. […] On the other hand, a resource may be generic in that as a concept it is well specified but not so specifically specified that it can only be represented by a single bit stream. In this case, other URIs may exist which identify a resource more specifically. These other URIs identify resources too, and there is a relationship of genericity between the generic and the relatively specific resource." (Berners-Lee, 1996,"Generic Resources")

> "This specification does not limit the scope of what might be a resource; rather, the term "resource" is used in a general sense for whatever might be identified by a URI. Familiar examples include an electronic document, an image, a source of information with a consistent purpose (e.g., "today's weather report for Los Angeles"), a service (e.g., an HTTP-to-SMS gateway), and a collection of other resources." (W3C Network Working Group, 2005, "Resource")

In both definitions the term resource is defined as both a very specific object, e.g. a picture, and a collection of other objects, which is a too broad definition to be useful for studying the use of single pages within a website. Within *Web Science* it is still debated what the definition of "resources" is (Halpin & Presutti, 2009). Concerning the Semantic Web one categorization of resource has been proposed by Halpin and Presutti (2009) where they distinguish between *Information Resource*, *Information Realization* and *Non-Information Resource*. The first, *Information Resource*, is an information object and "defined at a level of abstraction, independently from how it is concretely realized" (Halpin & Presutti, 2009, p. 529). In the terminology of FRBR it is on the level of *work* and *expression*, a non-materialized abstract form of objects. *Information Realization* is a concrete digital realization of the *Information Resource*, and the realization can be a whole book or a short message (a *manifestation* according to FRBR). *Non-Information Resources* represents objects that cannot "be digitized as a single digitally encoded message" (Halpin & Presutti, 2009, p. 529). "A single abstraction may have multiple realizations" (Halpin & Presutti, 2011, p. 279), but it is unclear if they mean that the realizations could be different parts of the resource. The three categories make sense within the semantic web where they have relations and attributes, but the proposed ontology by Halpin and Presutti (2009) does not solve the issue of entities of different scope outside the sematic web.

In the previous sections I have presented different definitions of the concepts object and resource. Due to the lack of a clear and useful terminology I use the term *object* in the thesis for the smallest

part of for example a digitized manuscript (an image of a page) or website (a web page), in the sense "relatively specific resource" in the Berners-Lee quote above and in the line with how a document is seen in the LIS perspective. The term *resource* is used at the higher level of aggregation, as a collection of digital objects (the whole digitized manuscript or a website), a "generic resource" in Berners-Lee's words. The two concepts are related to each other in different ways in the resource model and the object model (Figure 2.7). The models are used both in relation to usage and navigation strategies (Figure 4.4, Figure 4.5 and Figure 4.6), and in relation to findability (Figure 3.5).

2.4 Information seeking and retrieval

In order to study both the actions of the users (usage) and attributes of the CH resources (findability) a conceptual framework that incorporates the two dimensions in an equivalent manner. The integrated view on Information Seeking and Retrieval (IS&R) is an attempt to integrate the research in Interactive Information Retrieval (IIR) and Information Seeking (Ingwersen & Järvelin, 2005), both nested in Information behaviour (Wilson, 1999), as illustrated in Figure 2.8 where IIR is a part of information search behaviour. The IS&R framework is "based on a cognitive epistemological point of view and reflects an understanding of the information seeking and retrieval as a holistic *process* involving various cognitive actors in context" (Skov, 2009, p. 9). I have chosen to base the conceptual framework on the IS&R framework the central idea in the IS&R framework is the interaction between different cognitive actors, e.g. between user and information objects In the IS&R framework the cognitive actor or actors takes different roles, e.g. as creators of information objects, IT system designers or information seekers, and they act in a social, organizational and cultural context. "The actor stands out from the environment, so to speak, although still influenced by social interaction." (Ingwersen & Järvelin, 2005, p. 263), that is the explanation of the form of the right part of Figure 2.8.

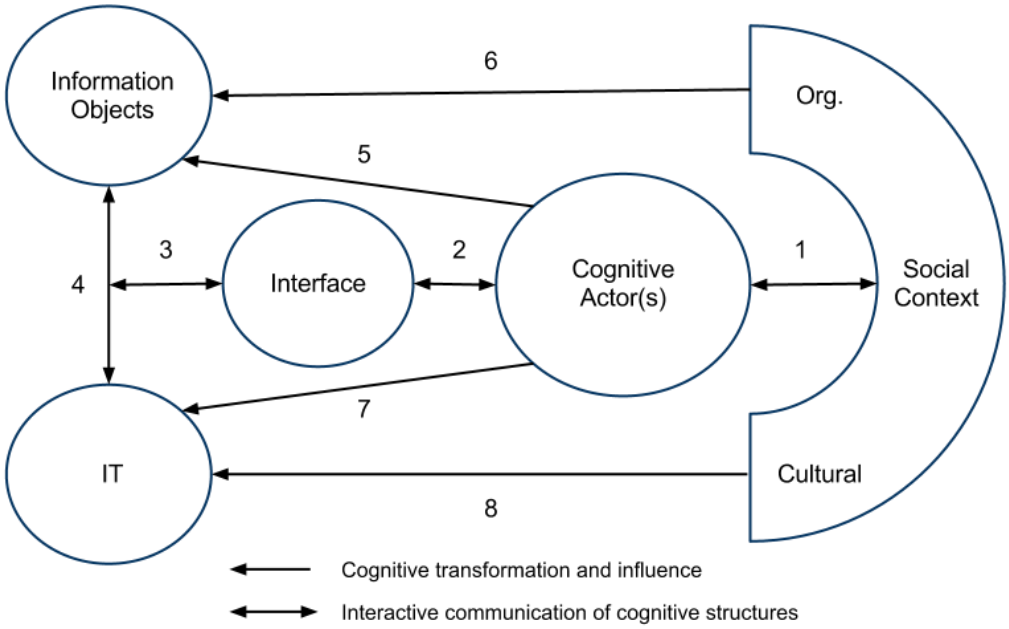


Figure 2.8. The IS&R framework, a generalized model of participating cognitive actor(s) in context in interactive information seeking, retrieval and behavioural processes, based on (Ingwersen & Järvelin, 2005, p. 261).

Arrows 1-4 in Figure 2.8 illustrates the processes of interaction and 5-8 illustrates different types of generation and transformation of cognition or cognitive influence. The framework has evolved from a cognitive model of IR interaction (Ingwersen, 1992, 1996) to more of a research framework (Ingwersen & Järvelin, 2005). The integrated IS&R research framework may be summarized into the three aspects: (1) the framework is general, the cognitive actors are not limited to information-seekers but they can be of all types, in all kinds of organizational, social and cultural contexts. (2) The framework is media-independent and integrates the social contexts with the cognitive IIR. (3) The framework offers guidance for multidimensional research designs (Xie, 2008, pp. 187-188).

The interactions in information seeking and retrieval between user and information system during search sessions is illustrated in Figure 2.8 as arrow number 2. From the user's point of view (arrow from *Cognitive actor(s)* to *Interface*) it is about the user's actions, information behaviour and search strategies. It is an interaction, but the studied aspect may be visualized as a flow from the user. From the system point of view (arrow from *Interface* to *Cognitive actor(s)*), the arrow illustrates the system feedback to the user's actions.

The main weakness of the IS&R framework in regard to the present project is the focus on one information system at the time. One of the main points behind the concept of findability is the complex relationships between the local information system (resource) and the web as a whole, the complexity is hard to capture in the present versions of the models. A web search engine can be seen as an interface to the numerous online web servers that are online. It could be seen as a

combination of interface, IT and information objects, but then is it hard to relate it to all the sites it has indexed.

Based on two versions of the IS&R framework, "Cognitive framework for scientific acquisition from nature" (Ingwersen & Järvelin, 2005, p. 273) and "Short-Term Interaction episode in IS&R carried out in a situation in context" (Ingwersen & Järvelin, 2005, p. 301) a modified version for the present study is presented below (Figure 2.9). The modified framework takes the complexity of the relations between information object, local information resource (web site) and the web into account. The framework describes interactions during a search session, the interactions during a limited amount of time, and therefore the influences over time are excluded.

In the IS&R framework arrow 2, the interaction between user and interface (or technology) is studied through RQ2 (about the usage of the resources). In RQ1 arrows 3 and 4 are studied, the structural relationship between the information objects and the technology (interface, system and the web). On the web the border between information and technology is indistinct.

> "It is less apparent what it means for people to interact with information. On one hand, since digital forms of information are so ubiquitous and require some kind of technology to facilitate human perception, people interact with the technology as an intermediary. This intermediate interaction with technology is tangible and necessary (but not sufficient) to accomplish information goals." (Marchionini, 2008, p. 170)

To make the distinction between objects, resource and the web clear on one side and the user on the other during information interaction Figure 2.8 is combined with Figure 2.7b. Figure 2.9 is a modified version of the IS&R framework design for the present research, designed with the weakness of the original IS&R model (Figure 2.8) and the tangible information interaction described by Marchionini in the quote above in mind. The numbers of the arrows in Figure 2.8 has been replaced by letters. A corresponds to 1. B, C and D correspond to 2, which highlights the complexity of the interactions on the web. A circle where the objects are embedded in a resource on the web has replaced the system aspects. During the information search process the user interacts with all parts of the circle (which is discussed in Chapter 4). The interaction within the system (arrows 3 and 4 in the original model) are different findability aspects (which are discussed in Chapter 3). An advantage of the IS&R framework model is that the information need or intention of the user is not an explicit part. Because it is implicit in the models the need of the user may advantageously be seen as a part of the interactions of the actors (arrows 1 and 2). Overall the IS&R framework models work well for mapping the different interactions during IS&R.

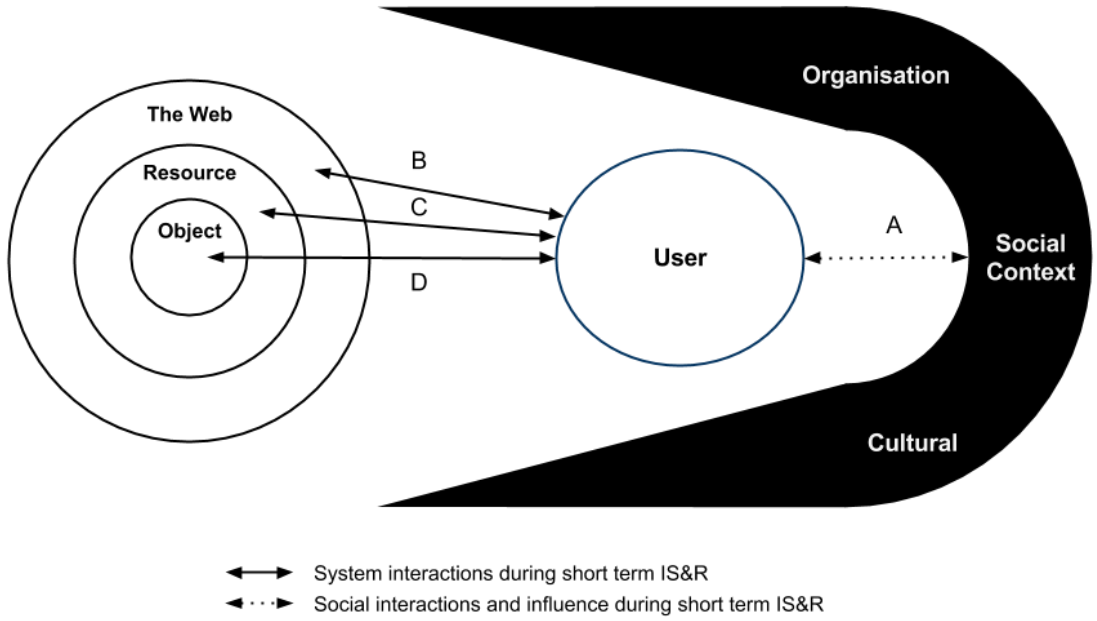


Figure 2.9. Web IS&R model, a modified version of Bates (2009a); Ingwersen and Järvelin (2005) IS&R model for the present study which takes the complexity of the web into account. The focus in the model is on short-term web interactions, i.e. the model does not include "cognitive transformation and influence" over time, which are included in Figure 2.8. The letters A-D describes different kinds of interactions.

In Figure 2.9 the interactions between the different cognitive actors in the present study has become clearer. The interactions between user and system are pictured as three arrows between the user and (B) the web; (C) a resource on the web; and, (D) an object within a resource (the three circles form Figure 2.7b). The object is embedded within a resource, and the resource is accessible on the web. The user interacts with a specific resource or with the web to get to a resource, both acts as interfaces of the information objects. Objects might be accessed directly on the web, but they are always part of a resource (even if the resource just consists of one object, e.g. a web site with one web page). These interactions are the usage and navigation studied in the first research question (RQ1). The distinction between B, C and D might be seen as artificial, but it is crucial when studying search behaviour. Pharo makes the same distinction in the SST method schema; search situations and search transitions (Pharo, 2002). One aspect is that the focus of an exploratory search gets narrower; browsing is replaced with search to a larger degree (White & Roth, 2009). Another aspect emphasized in the web IS&R model is that the web navigation behaviour is replaced with other types of interaction when the user has arrived to a resource; in the resource the web navigation is transformed into site navigation as the two types of information systems has different attributes.

The system aspects, the findability of the objects, are not labelled in Figure 2.9 but the object, the resource and the web all have relations with each other, and these settings and links constitutes the information environment the user interacts with. In the original model (Figure 2.8) the relationships are illustrated differently because the model is created for information retrieval

systems where the basis for interaction is search by queries. The interface interacts with the information objects and IT in the IR system. Search by query is a part of the modified version of the model (Figure 2.9), but it is just one of several modes of navigation or interaction covered by the model.

I have chosen to base the conceptual framework on the IS&R framework as described in the section because the central idea in the IS&R framework is the interaction between different cognitive actors, e.g. between user and information objects. This means that I view the interaction between the cognitive actors in a certain manner. Over time the cognitive actors influence each other (Figure 2.8). I study short sessions in the interactions between the user on one side and the objects, the resource and the web on the other as illustrated in Figure 2.9. The crucial point here is the interaction, not just the actions of the users, but the actions of the user and actions of the system in form of feedback and movements in the information space forms a whole. The system in form of objects within resources on the web is the counterpart to the user.

2.5 The information search process

The starting point of the thesis is that the navigation and search on the web is not only dependent on user characteristics such as information skills, personality and motivation, and her information needs, but also the context of the search, i.e. the interaction with information. The overall information system is in this case the web and no matter where or how a person search on the web she is in a context, in a resource with a specific content.

The web is a large and complex information system consisting of numerous local information systems connected in a network, the Internet. Therefore, navigation and search the web is more complex than it is in a single information system, such as a website or database. In information systems, or information spaces, the users interact with different functions and respond to information that they encounter. Two information seeking theories or models take the users' adaption to the information environment into account: *Berrypicking* and *Information Foraging Theory*[4] (White & Roth, 2009). Bates uses berrypicking as a metaphor for a real-life search behaviour, and in contrast to the classic IR-model where queries are matched against document surrogates (Bates, 1989). The original description of berrypicking states:

> "In real-life searches in manual sources, end users may begin with just one feature of a broader topic, or just one relevant reference, and move through a variety of sources. Each new piece of information they encounter gives them new ideas and directions to follow and, consequently, a new conception of the query. At each stage they are not just modifying the search terms used in order to get a better match for a single query. Rather

[4] Also called *Optimal Foraging Theory* within LIS (Sandstrom, 1994, 1999).

the query itself (as well as the search terms used) is continually shifting, in part or whole. This type of search is here called an evolving search.

Furthermore, at each stage, with each different conception of the query, the user may identify useful information and references. In other words, the query is satisfied not by a single final retrieved set, but by a series of selections of individual references and bits of information at each stage of the ever-modifying search. A bit-at-a-time retrieval of this sort is here called berrypicking. This term is used by analogy to picking huckleberries or blueberries in the forest. The berries are scattered on the bushes; they do not come in bunches. One must pick them one at a time." (Bates, 1989, "II. A "Berrypicking" model of information retrieval")

The berrypicking adapts to the environment and exploits the possibilities as they emerge as illustrated in Figure 2.10. The query, the expressed information need, is dynamically influenced by the information objects looked at and through thought during the information searching. The berrypicking-path is a move from object to object; moves within and between the zones in the information source horizon which forms the information pathway (Savolainen, 2008).

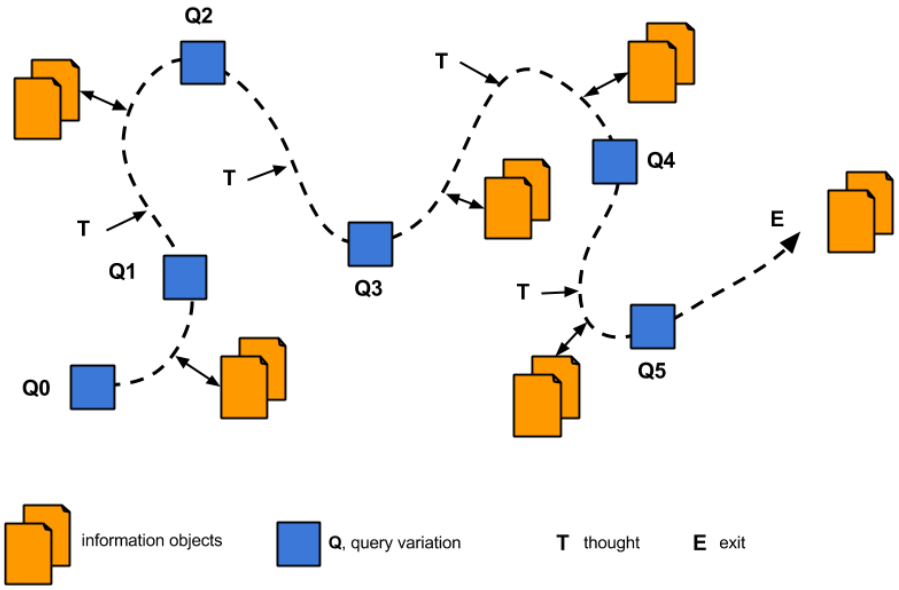


Figure 2.10. Bates Berrypicking model of search (Bates, 1989). The model is slightly modified, the arrows between the user on the berrypicking-path and the viewed information objects has been made bidirectional because the object might have impact on the path and the queries.

Information Foraging Theory has a similar approach, it "is an approach to understanding how strategies and technologies for information seeking, gathering, and consumption are adapted to the flux of information in the environment" (Pirolli & Card, 1999, p. 1). In Foraging Theory the foraging behaviour is seen as strategic, not random actions. The foraging models are based on the assumption that the forager (searcher) has limited energy and wants to maximise the intake per energy unit spent (Hantula, 2010). The information foraging takes place in a dynamic ecology

were particular environmental conditions are central, e.g. *information scent* and *information patches*, and the theory "provides information seeking researchers with a way to examine users' goals, their decision-making processes, and adaptation to the information environment" (White & Roth, 2009, p. 28). Explorative search, berrypicking and information foraging are similar in several ways, but a central difference is that in information foraging "the prey" is known in advance and the goal in the theory is the maximisation of utility, according to White and Roth (2009). It ought to be possible to combine the theories despite their differences as they analyse the information search behaviour on different levels of analysis, macro and micro behaviour within a search session. In Chapter 8 navigation behaviour of the users is analysed and in Chapter 9 the intention and context of their visit in the CH resources.

2.5.1 Query-dependent and -independent aspects

One dimension in information searching is the user in relation to the information system. On the user-side, we talk about information needs and information search skills, and on the system-side the embedded information and the complexity of the web. Another dimension is the question of which parts that are query-independent in information searching. The user's current information need creates a query-dependent context for the search. However, regardless of what information is needed, much of the information on the web is not relevant. An example is how Google handles a normal search. When a query is sent to Google the search engine first calculates what content that is most relevant based on keywords. Then Google determines the ranking on the relevant web pages based on the relevance calculations and PageRank of the web pages. PageRank is Google's measure of the prestige of a webpage on the web, i.e. which other pages has links to the page, the prestige of the pages linking to the page and how many links are there to the page. PageRank is a quantitative measure based on link analysis and is completely independent of the search terms used (Brin & Page, 1998). The ranking is thus a mix of query-dependent and query-independent aspects.

Gerjets and Hellenthal-Schorr describe information search skills and its partial competences independent of both the medium and the user's goals (Gerjets & Hellenthal-Schorr, 2008, p. 696):

> "These sub-competencies are usually conceived as being independent of the specific medium used and of users' goals in the context of media utilization."

The core competencies in information searching on the web enable information searching, but are general. Information search skills are separate from the information need and search subject. Subject knowledge relevant to the information need is of course in play, but research suggests that knowledge in the subject area (domain knowledge) only affects the search behaviour, not the search efficiency (Zhang et al., 2005). This means that on the user side there are both query dependent and query-independant elements.

The same division between the query-dependent and query-independent elements can be made on the information system side. The content of the information objects and their representation in

the form of metadata is the information the user match against his information need, and can therefore be viewed as a query-dependent aspect. Some of the representations of the objects, the content of the metadata, must match some part of the information needs to be perceived as relevant by the user. An example of a query-dependent is the snippets in Google that change for the same object depending on the query. In *Information Foraging Theory* the concept of *information scent* is used when the user judges the clues to or representation of potentially relevant items. By "the scent" from the clues, for example link texts on a web page, the user will select the one most likely to lead to the goal, i.e. who has the strongest scent (Pirolli, 2007). If the smell fades out, the back button in the browser is often used to return to a previously visited page and from there follow another scent trail, or instead start a new search. The representation of the object is the aboutness, the topic of the object.

How accessible the objects are technically and structurally is a query-independent aspect. How and where the information objects and their representations are visible in the system and on the web is also important from a query-dependent approach. The amount of all metadata, not the topicality of the metadata, creates a "target area" for information seekers when searching in search engines. A central aspect of the web is that objects have their own unique web addresses, URLs, so they are accessible directly and does not require a further search in a database.

2.6 The Webometric perspective

In addition to the perspectives of ELIS and Information Seeking and Retrieval, the study can also be seen as *Informetrics*, and more specific *Webometrics.* Informetrics can be defined as:

> "Informetrics is the study of the quantitative aspects of information in any form, not just records or bibliographies, and in any social group, not just scientists. Thus it looks at the quantitative aspects of informal or spoken communication, as well as recorded, and of information needs and uses of the disadvantaged, not just the intellectual elite. […] Although in practice the scope of informetrics is very broad, in the past, bibliometricians and scientometrics have concentrated their studies of mathematical models and measures in a few well defined areas" (Tague-Sutcliffe, 1992, p. 1)

Tague-Sutcliffe lists different areas of study including characteristics of publication sources and use of recorded information. She is at the same time aware of other possible areas of study, such as definition and measure of information which has not been seen as a part of the informetric tradition (Tague-Sutcliffe, 1992). One can argue for that findability analysis is a part of informetrics as well as the study of the use of web resources is a part of informetrics. The part of informetrics focused on web aspects is called *Webometrics* and is defined by Björneborn as:

> "The study of the quantitative aspects of the construction and use of information resources, structures and technologies on the Web drawing on bibliometric and informetric approaches." (Björneborn, 2004, p. 12)

Ingwersen and Björneborn state that the objectives of common webometric analysis are studies of *selected web spaces*, *web indicators* and *human actor-web interaction* (Ingwersen & Björneborn, 2004, p. 347). The present study has aspects from all the three objectives. The thesis examines three Danish cultural heritage resources with digitized material online (*selected web spaces*) in several different ways. Through the log files of the resources the users' navigation strategies are explored (*human actor-web interaction*). The findability of the resources and their objects are measured (*web indicators*) and put into relation to the navigation patterns. According to Björneborn and Ingwersen there are four main areas in webometric research:

> "This definition thus covers quantitative aspects of both the construction side and the usage side of the Web embracing four main areas of present webometric research: (1) Web page content analysis; (2) Web link structure analysis; (3) Web usage analysis (including log files of users' searching and browsing behavior); (4) Web technology analysis (including search engine performance)." (Björneborn & Ingwersen, 2004, p. 1217)

The four areas in webometric research, in the quote above, studies different types of data. In the thesis three of the four types will be studied: content (1), structure (2), and usage (3) (Björneborn & Ingwersen, 2004). The fourth, Web technology analysis is not studied in an independent manner, within the present framework it might be seen as the combination of content and structure of the resources. I view the three types as different levels of interaction within the unspecified interaction between the user and the interface of the system (arrow 2 in Figure 2.8 and arrows B, C and D in Figure 2.9). Links and other parts of the html and http protocols are the *structural level*, along with other internet infrastructure aspects. The content, texts, pictures, etc. is built upon the structures in the structural level and forms a *content level*. The usage of the web, the navigation, searching, reading, playing, etc. is an utilization of the content (read, play) and structure (navigate, search) and is here viewed as an *usage level*. The three levels will be a way of relating different kinds of methods and data together into a coherent whole. The object, resource and web layers can be combined with the levels of analysis as well as time and thereby constitutes a three dimensional model with the purpose of relating phenomenon at different levels to each other (the URI model in Figure 2.11). The three webometric data types are important components in the URI model in the form of levels because the interaction between user and resource can be differentiated and studied individually.

2.7 The User-Resource Interaction model

Based on the research discussed in the previous Sections 2.3-2.6 I have created a fundamental conceptual model for user-resource interaction. The two conceptual pairs of user/resource and query dependent/query independent aspects have been combined with the three webometric levels in order to create a tool for both research design and analysis of empirical data. In the model, Figure 2.11, the terms information need, search skills, representation of information object, and

findability shall be seen in the context of the interaction between users and systems during a search session. The model has traits from both Belkin's episode model of interaction with texts (Belkin, 1996) and from Saracevic's stratified model (Saracevic, 1996, 1997), especially the notion of the meeting between the user and the information system with objects, and time as a dimension in the model. The User-Resource Interaction (URI) model in Figure 2.11 depicts the status at a specific time during search, and a search session consists of a large number of snapshots similar to Belkin's ideas (Belkin, 1996). The user (I) interacts (G) with the resources and objects on the web (A) based on both query-dependent content aspects (B and F) and query-independent structural aspects (D and H). The third axis (J) in the model high lights time as a factor and that the variables changes for each interaction during the search process (E).

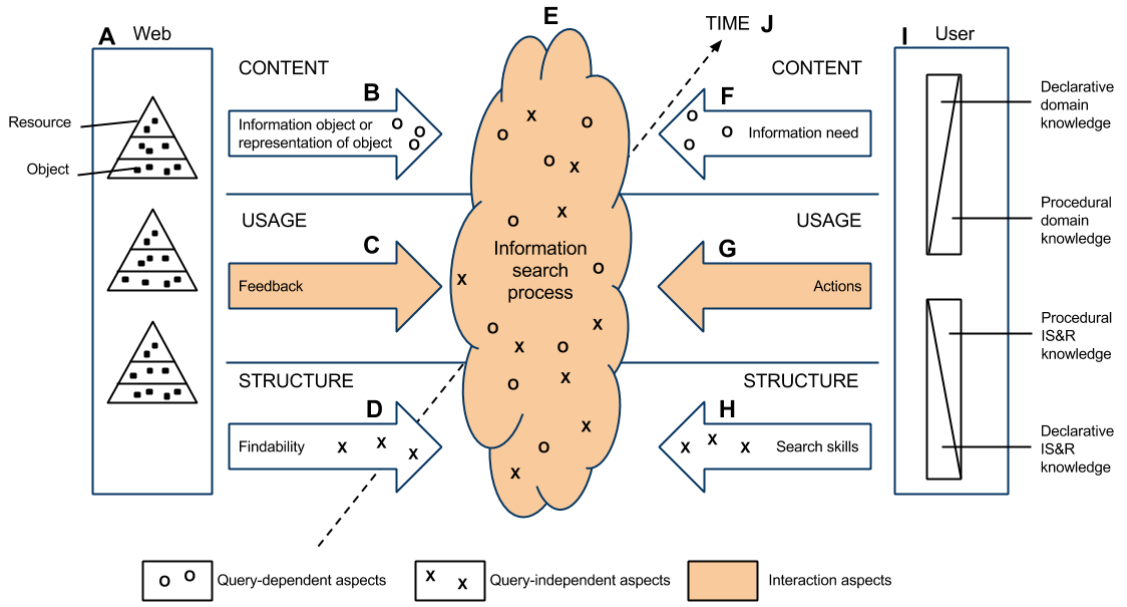


Figure 2.11. The User-Resource Interaction model (URI model) of the interaction between user and the web (including objects and resources) during searching. The elements in the model are marked A-J, which are used as references to specific parts of the model in the text in all chapters. The query-dependent aspects (marked with O) and query-independent aspects (marked with X) are mixed in the information search process as the actions of the user depends on aspects of both types.

On the right side of the model is the user who initiates an information search process with her information need and search skills, as well as domain and IS&R knowledge (Ingwersen & Järvelin, 2005). They correspond to the representation and the findability of the information objects. In the information search process, the user processes and responds to the feedback from the web and local information systems which is provided as a reaction to the user's actions. In the upper part of the model are the elements of the search process which deals with information needs, i.e. queries and the content and representation of the information objects, which are the query-dependent parts. In the lower part are the query-independent, more structural parts i.e. search skills and findability. In the model the interaction aspects are shaded.

Although the arrows point at each other, they should not be seen as in opposition to each other. On the contrary, they complement each other at each level in the model; they are input variables in the information search process. For instance, users with a high level of search skills are more likely to find objects with low findability than users with a low level of search skills, under the assumption that both objects are perceived as relevant.

In the model in Figure 2.11 the length of the arrows might illustrate the degree of search skills and findability, and the clarity and specificity in the information need and the representation of the content of the objects. It illustrates the assumption that a low level of search skills means that only items with high findability are found. The user has fewer search strategies to use and less efficiency in all of the sub-processes in the information search process. Likewise, objects with a low findability demands a higher level of search skills from the users so that they can be found or reached.

The model makes no claim to describe the whole complexity of navigation or search on the Web. The point of the model is that it divides the input variables into two groups: query-dependent and query-independent, and that it separates the user from the concept of findability. The model does not take into account the structural aspects of social, affective or personal nature, but they can be seen as part of the search skills at the moment of search. Nor is information need specific aspects taken into account, such as motivation and ambition. They can be seen as aspects of a multidimensional information need.

The model and more specifically the information search process-cloud in the middle of the model does not try to explain the process of information seeking. There are numerous attempts to explain or model the search process, e.g. Wilson (1981, 1999), Kuhlthau (1991), Marchionini (1995) and Xie (2008). The models in Figure 2.7 and Figure 2.11 are enhancements of the activity represented by arrow 2 in Figure 2.8, the interaction between actor (user) and interface (as a representative of the whole system). In the model X and O signify that at a given point in time the information search process consists of a combination of the elements, together with the actions of the user and the feedback from the system (the resource), coloured in the model.

If the model is seen in a more general context, many people work to bridge the gap between the lack of information literacy and low findability. Information architects work on navigation and search on websites and search engine optimizers struggle to optimize web pages for as high ranking as possible in search engines. Librarians and other information specialists act as "compensators" and intermediaries between collections with low findability and the users' insufficient information search skills. They also are increasingly tasked to train the users' search skills by teaching and other activities (Chevillotte, 2010).

The consequences of low search skills and low findability are from a larger perspective challenging for both the individuals and the society. There are enormous amounts of information on the web (which continues to grow), while some actors pay for increased findability of their information (requires resources). This means that other non-optimized information such as alternative views or digitized cultural heritage may require more effort and better search skills to

be found by users. This can be seen as an existential problem, as Peter Morville puts it: "What we find changes who we become." (Morville, 2005, front page).

In the URI model (Figure 2.11) the feedback arrow (C) could be put into a parenthesis, as explicit feedback is only generated in some information systems, e.g. search engines and their system-based relevance (Borlund, 2000). All other "feedback" is in the form of new web pages with new content and structure, which is not an action in itself but a movement in the information space.

The URI model is the conceptual framework in condensed form. In the model the user interacts with information objects over time during a search session. The interaction is divided into the content, usage and structure levels. The upper part of the model covers query-dependent aspects and the lower part of the model query-independent levels. All the aspects are a part of the information search process. The present version of the URI model is focused on interactions with CH resources on the web but the model can be used for all types of information interactions regardless of medium with small changes.

2.8 Chapter summary

In this chapter the IS&R framework is used as a foundation for the conceptual framework, and both usage and findability are seen as parts of the interactions between the user and the system (Figure 2.8). As the original IS&R framework does not take the complexity into account when the users moves between the web and specific web sites, a modified version of the model was developed (Figure 2.9) where it is possible to make a distinction between the two types of navigation. The web IS&R model also includes the circular object model from Figure 2.7b, where objects are seen as embedded in a resource on the web. With the URI model in Figure 2.11 I have made a model that explicit builds on the IS&R model (Ingwersen & Järvelin, 2005) and has traits from Berrypricking (Bates, 1989), ASK (Belkin, 1996) and Saracevic's stratified model (Saracevic, 1997). The model of user-resource interaction (URI) focuses on the interation during the information search process, but also deals to a large degree with information seeking and web science/webometrics. In the URI model the system side is represented by resources on the web (Figure 2.7a).

Several important distinctions were made in the chapter. One distinction between different types of analysis is found in the webometric research tradition: content, structure and usage (together with the analysis of web technology). The three levels of data are used to divide the information search process into levels, both on the user side and the system side (expressed in Figure 2.11). Another important distinction is the one between object and resource. A resource is a collection of objects, a web site or sub site with a clear focus. The relationship between the two concepts is displayed in Figure 2.7. The conceptual framework expressed in Figure 2.11 is a way to relate usage and findability to one another. By using the framework it is possible to investigate and discuss interactions between users and information systems in new systematic ways. The actions

of the users can be compared with both the topicality and the structure of the content, as well as the feedback from the system.

The ELIS framework is a supplement to the conceptual framework expressed in Figure 2.11 in that the users and the information search process are seen in the context of everyday life information practice. ELIS is partly used as an analytic tool for interpretation of the results of the web survey and the log findings.

The developed conceptual framework and the URI model in Figure 2.11 is the foundation of the whole thesis. In the next three chapters theoretical aspects are covered. In Chapter 3 the information objects, information resources and information system are in focus, especially the findability aspects (the left side of the URI model). In Chapter 4 the focus is on the right side of the URI model, the user and her actions, the actual usage of the CH resources. Chapter 5 covers the research design and methods as a bridge to the empirical part of the thesis (Chapter 6 to 11). In the empirical part of the thesis the methods are applied according to the framework (Chapter 2), the site structure analysis (Chapter 6), the findability analysis (Chapter 7), the transaction log analysis (Chapter 8) and the survey data (Chapter 9) integrates different aspects of the framework in the analysis.

3 Cultural heritage objects and their findability

> "Accessible content is findable content" (Walter, 2008, p. 29)

This chapter addresses the structural aspects, with a focus on the site structure based on content and function, and findability of both the resources and the objects. The system side aspects (A-D in Figure 2.11) are the interactions or reaction from the system on the users' actions. Findability is the most important system aspect in the present study. The major question in this chapter is: *What aspects are important for measuring the findability of a web object?* Findability is a relatively new concept which has only been used or discussed briefly in previous research. I will define the concept and divide it into six important aspects, which will be measured or evaluated in the empirical parts or the thesis.

First is the question about how the internal structure of resources can be divided into levels address. Dividing the resources into different levels according to the resource model (Figure 2.7a) is crucial both for the findability analysis and the study of the log files. The user aspects of the user-system interactions are covered in Chapter 4. In the later part of the chapter cultural heritage resources are discussed and the studied Danish web resources are presented as well as the criteria for the selection of the CH resources.

3.1 Structure and content

3.1.1 The challenge of the environmental context

The total information system as an environmental context of users' actions can be seen in two different ways in relation to information searching on the web. Either the focus is on the resource (local system) and it is seen as the primarily unit, which contains objects and is linked together with other resources on the web. Or the web is seen as superior to the resource and the objects of the resource are regarded as a part of the contents of the web together with all other available objects. I have chosen both perspectives as it is not possible to choose just one of the perspectives if the point of departure is the users' information searching. Depending on search strategy the contents of the web is handled in different ways. In direct search the single objects is the primarily goal, e.g. in a topical search in a search engine, while in an indirect search in a search engine or during browsing the contents of the resource as a unit are more important. When the users have found the resource, the search continues among its objects. This dual perspective leads to two kinds of findability, *external* and *internal* findability.

The left side of the URI model (Figure 2.11) represents the resources. The upper arrow (B) represents the content, and the lower arrow (D) the structure of the content and the system. Based

on this model the relevance evaluations (in both human and system) take place in the upper part of the model, in the query dependent part, and IR problems such as the language problem (Petras, 2006) are content based. The query independent lower part is much more indirect and subtle. The structure of the system and its content determine the likelihood of what object the user will be exposed to and evaluate the relevance on.

In Web IR a well-known example of an algorithm that works only as a query independent measurement of the structure is the PageRank algorithm (Page et al., 1999). Another example of an algorithm competing with PageRank is HITS (Kleinberg, 1999). In HITS a set of relevant pages is chosen based on the query, the links between the pages in the set is then analysed and two kinds of pages are identified: hubs and authorities. Hubs are pages which link to several of the authority pages, pages that are identified as highly relevant to the query. Contrary to PageRank HITS works on both the query dependent level and the independent level, the link structure is used to identify authoritative pages on the topic of the initial query (within the query dependent set). PageRank is used as a part of the ranking system in Google, where it is combined with traditional IR text relevance calculations. Both PageRank and HITS has been developed to promote authority and thereby enhance the retrieval, which traditionally mostly focuses on the query dependent aspects. The two algorithms highlights the importance of the structural aspects of the information space, which in the thesis is studied through the concept of findability.

3.1.2 Linking

> "Links connecting pages are a key component of the Web. Links are powerful navigational aid for people browsing the web, but they also help search engines understand the relationships between the pages." (Croft et al., 2010, p. 106)

Links on the Web can be looked at from several perspectives. Björneborn has created a terminology for link relations between web nodes (2004).

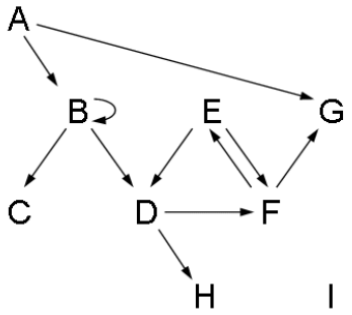


Figure 3.1. Basic link relations (Björneborn, 2004, p. 16).

The relations are described below, in a simplified version of Björneborn (2004).

- B has an *inlink* from A; A is *inlinking*
- B has an *outlink* to C; C is *outlinked*

- B has a *selflink*
- A has no *inlinks*
- C has no *outlinks*
- I has neither inlinks nor outlinks; I is *isolated*
- E and F have *reciprocal links*
- A has a *transversal* outlink to G: functioning as a shortcut
- H is *reachable* from A by a *link path*

In studies of the web it may be useful to visualize relations between different units of analysis (Björneborn & Ingwersen, 2004). One model is the *Alternative Document Model* (ADM). It is a model of what to count and how to group the web pages. According to Thelwall and colleagues, often the most appropriate unit of measure will be either the Web page or Web site. There are four versions of ADM: *page*, *directory*, *site* and *domain*; and the pages are seen as part of the entities according to the ADM version in use, e.g. in site ADM all pages within the site is treated as one entity (Thelwall et al., 2005).

In the web node framework Björneborn has illustrated the basic building blocks with geometrical figures: *quadrangles* for Web pages, *diagonal lines* for directories within sites, *circles* for sites, and *triangles* for country or generic top level domains. Additional borderlines in the geometrical figures means it is a sub-level entity, e.g. a sub site (Björneborn, 2004).

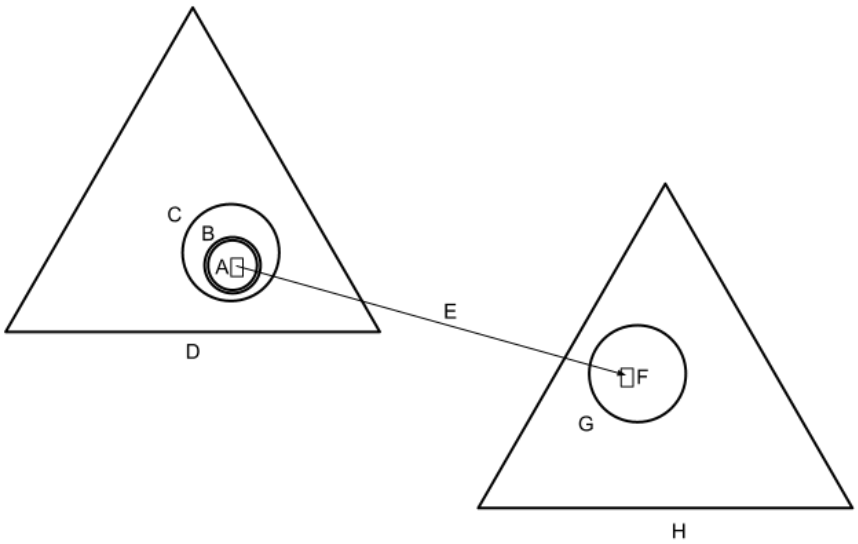


Figure 3.2. Web node diagram with page level links based on Björneborn (2004, pp. 19-21).

In Figure 3.2: Page A in sub-site B (e.g. da.wikipedia.org) a part of site C (wikipedia.org), within the top level domain D (.org) has an outlink E to page F. Page F has an inlink E and is placed in site G (adl.dk), within the top level domain H (.dk).

The web node framework is based on the structure of the URL with an emphasis on site and domain, but more important also on the file structure on the web server. The file structure has

become a less important measure of how the navigation of the site works. In many sites the content (the pages) is stored in a database and is dynamically presented as pages when requested. An example is the following URL for a work of art in Art Index Denmark where the requested "page" has the id 42250 (the question mark is an indicator of dynamic content): https://www.kulturarv.dk/kid/VisVaerk.do?vaerkId=42250.

The web node framework is central for link analysis between pages, sites and domains, but it is not a way to picture or map the navigation structure of sites or sub-sites. The internal information architecture may be totally separated from the file structure, which is increasingly the case, as content objects are stored in databases instead of as html files in hierachly organised folders. Thereby it is more interesting to divide the resources into levels (as in Section 2.3) than based on the URL structure. The linking is crucial for findability, both for the objects to be reachable and to gain prestige (authority) on the web, which is discussed in Sections 3.2 and 3.3, and will be used in Sections 8.2.1 and 8.2.3 on findings of the log analysis.

3.1.3 Content analysis

There are two main ways of looking at content (information) on the web:

A. Content are divided into information objects and representation of objects, where the objects are the "real" content and the representations are information about content (metadata).
B. All information objects are content; no separation is made between objects and representations.

The two ways are immanent in for example IS&R-research in different ways. In the thesis a division is made between object and resource, which is in line with perspective A. Another example is Pharo's SST method, where the search interactions are divided into the *search transition* (towards interaction with information) and the *search situation* (interaction with information) (Pharo, 2002).

Another categorization of web content is that of Almind, as cited in Almind and Ingwersen (1997):

- Personal home page (represent an individual)
- Institutional/organizational home page (represent an organization)
- Subject defined/*ad hoc* home page (represent a subject)
- Pointer document/index page (make hyperlinks available)
- Resources (make data available)

A third variant is the categorization of Haas and Grams. It is consisting of seven page sub classes (Haas & Grams, 2000):

- Organizational

- Documentation
- Text
- Home page
- Multimedia
- Tool
- Database entry

The categorizations above are of type B and they are closer to genre than type A. Björneborn studied interlinking within an academic web space on a genre level and found 17 genre classes divided into two categories: personal and institutional (2011b).

The content are in some ways closely connected to structural aspects, especially the part of findability called "attributes of the object". The file format of the object has implications on both content and structural levels. Text in plain html has some characteristics affecting both levels; a sound file has other characteristics. The impact of the file type on the content level is determined by categories used in the content analysis.

In the thesis the content are of similar type. It is made up of digitized cultural heritage, with the aim to reach the general public, the citizens. There are problems on how to handle metadata for information objects other than text. The metadata about the content of a sound file could be considered as a part of the object, or the metadata could be seen as a representation pointing towards the object. Or, in line with view B above, as independent information object (with a relation to the sound file). In the present thesis this problem is solved by considering objects of either type (containing digitalized CH or just metadata) as objects as long as they have an unique URL. Objects, different types of content, can be placed in different parts of the model as illustrated in Figure 3.3 (se also Figure 3.5 where the structural findability aspects are displayed).

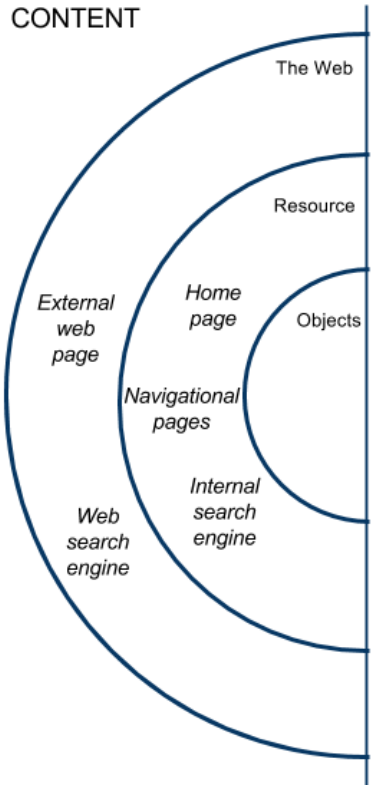


Figure 3.3. Illustration of potential kinds of content in the object model (Figure 2.7b).

To make a division between resource and object is not enough as stated in Chapter 2, a resource is a collection of objects. The triangular-shaped resource model (Figure 2.7a) is used to make distinctions between different levels and types of objects based on the function of the objects in regards to both the usage analysis and the findability analysis in line with the distinction discussed in Section 3.1.2.

3.2 Findability

To be found is a goal for everything published on the web. For public services there are three main reasons for working on their findability: (1) It is a good investment to be found by the users, otherwise the digitalization has been in vain or at least less effective. (2) Reaching the users is good service. And (3), it is good for the democracy that the citizens can find what they need when they need it (Høgenhaven & Lundberg Andreasen, 2011).

3.2.1 Defining Web findability

Findability is a complex concept. The Wikipedia defines findability that follows:

> "Findability is a term for the ease with which information contained on a website can be found, both from outside the website (using search engines and the like) and by users already on the website. Although findability has relevance outside the World Wide Web, it is usually used in the context of the web." (Wikipedia, 2013)

The Wikipedia definition relies heavily on Peter Morville's definition. He defines findability in *Ambient findability* as follows (Morville, 2005, p. 4):

a) The quality of being locatable and navigable.
b) The degree to which a particular object is easy to find or locate.
c) The degree to which a system or environment supports navigation and retrieval.

According to Moville's definition findability operates on different levels, both on object level (b) and on resource (system) level (c). He also talks about quality of and degree of findability, which means that findability are at some scale. Information on the web is published in some information resource (e.g. content management system, blog or database) available on the web. Because of the complexity of web publication findability is constantly changing and hard to calculate. But it is the most important aspect concerning information on the web in this era of search engine use (Halavais, 2009).

In the research literature there is no clear definition of the findability. The concept is related to the concepts retrievability and searchability. Retrievability used in IR research, and measures how likely it is for a document to be found with a given set of queries. Sometimes the terms retrievability and findability are used as synonyms. Bashir and Rauber writes:

> "In recent years measurement concepts like document "retrievability", "searchability" and "findability" emerged [...]. These concepts measure, how retrievable each individual document is in the retrieval system, i.e. how likely it is that a document can be found at all given a specific set of queries" (Bashir & Rauber, 2009, p. 753)

While Azzopardi and Bache makes the following distinction between findability and retrievability:

> "The accessibility of information in a collection given a system has been considered from two points of view, the system side i.e. retrievability […] and the user side findability […]. Retrievability measures provide an indication of how easily a document could be retrieved using a given IR system, while findability measures provide an indication of how easily a document can be found by a user with the IR system." (Azzopardi & Bache, 2010, p. 889)

In the article Azzopardi and Bache do not discuss the difference further. Shall findability be seen as the user side of retrievability? If retrievability is the measurable system perspective along with concepts such as precision and recall, does findability include the user's information needs and level of search skills?

Azzopardi and Vinay points out the requirements on different levels that must be met before findability can be achieved in an information retrieval system like a search engine. The document must be indexed to be findable, that is the first requirement. Then, the document can be found by searching in the system. The second level consists of several factors, the matching system in the retrieval system, the user's ability to convert their information needs into an appropriate query for the system (search skills), and the user's motivation to look through the many documents or hit lists. Azzopardi and Vinay concludes that if you combine these factors it means that some documents are more easily accessible than other documents (Azzopardi & Vinay, 2008). Here is the user and her information skills an explicit part of the findability.

Findability is sometimes described as an active process, a process from the publishers side who take actions to make the information findable, but also from the user's side who actively searches for the information (Lutze, 2009). But from the user's side it is about searching and finding. Findability in this sense is about the user's information searching. The concept which connects the searching of the user with the findability of the document is relevance. Perceived relevance emanates from the user's information need and is constantly matching need and document throughout the information search process, as Ingwersen states: "relevance is ultimately a value of pragmatic nature, linked to the individual user's problem space and state of knowledge" (Ingwersen, 1992, p. 54).

The professional field of *Search Engine Optimization* (SEO) works with some aspects of findability to promote the ranking of web pages in search engines. In SEO both on-site and off-site aspects are regarded in the process. The knowledge in the field is mainly based on experience, trial-and-error, due to the secrecy of the search engines algorithms and their constant change. Another professional field dealing with findability is *Information Architecture* (IA). In IA there is a focus on findability and navigation within a website and the field can be seen as a part of the practice of web design and usability.

In the book *Building findable web sites* findability is described as the common thread in all aspects of web publishing, from copy writing, design and information architecture to development, search engine optimization and marketing, plus accessibility and usability (Walter, 2008). Findability is illustrated as a flower with many petals. In this web design sense findability is a main concern for information creators, owners and publishers, not users. This is supported by the following conclusion:

> "It should be noted that the information gap surfaced as a critical factor for organizations but not for users. One possible explanation is that the general public views the entire World Wide Web as the information source rather than a particular web site. Hence, if the information sought is not available from a specific web site, visitors move on to the next one." (Angelov et al., 2010, p. "Conclusions")

The Wikipedia definition of findability is in line with Morville's definition; findability is defined as a user free and query independent concept because there is no concept of findability beyond retrievability from a web perspective. On the user side, we have different notions of knowledge,

skills, competencies and needs, we do not need another concept that includes several already elusive aspects of human information behaviour.

Based on Morville's definition and adjusted for collections of digitalized cultural heritage on the web, my definition of web findability is:

The degree to which a particular object or resource is easy to find or locate (within a web site or on the web). It is depending on

a) characteristics of the object,
b) design and settings of local information system (web site) wherein the object is published or stored,
c) The prestige of the object and of the web site (resource) in form of inlinks.

Findability is in some senses an extended version of retrievability which takes the whole web into account, not just a local information system. But as it takes both the resource and the web into account it is a much more complex concept. Findability is used to measure and compare how findable different items are. How findable an item is for an individual user depends on both the degree of findability of the object and on the user's level of search skills. Findability, however, describes a group of characteristics of the object that can be found.

3.2.2 How is something findable on the web?

To be findable documents on the web needs to have certain characteristics. The characteristics are different depending on which of Levene's three information searching strategies that are used (Levene, 2010; Nachmias & Gilad, 2002). The strategies are discussed in Section 4.2.3.

The strategies are often used in different combinations. First a search engine might be used to locate the top page of a website, and then the user navigates within the site to find the relevant object (which might not be indexed by the search engine). This indirect way of searching requires some knowledge on the topic of the user, which is not required in the same degree when using a directory or search engine. Since the 1990s the general search engines has become the most important type of search services on the web. They have become the gatekeepers of the web, and can even be called the web dragons (Witten et al., 2006).

Everything is not in the indexes of the search engines. Two different concepts are used for the information not found in the search engines: *the invisible web* and *the deep web*. The invisible web is the part of the web not indexed by a specific search engine, and thereby the information is invisible in that search engine – the invisibility is specific to a certain search engine. The deep web is the opposite of the shallow, easily indexable web made up of web pages in HTML. It is information stored in databases and will only be presented dynamically when searched for; the search engines cannot execute search queries and the information is buried too deep for indexing (Sherman & Price, 2001). The information can also be in a format that is hard or impossible to index, like pictures, videos, programs or zipped files (M. K. Bergman, 2001). In recent years the

two concepts has become less distinct since the search engines has become more technical advanced and stores larger amount of data in their indexes (Fransson, 2007).

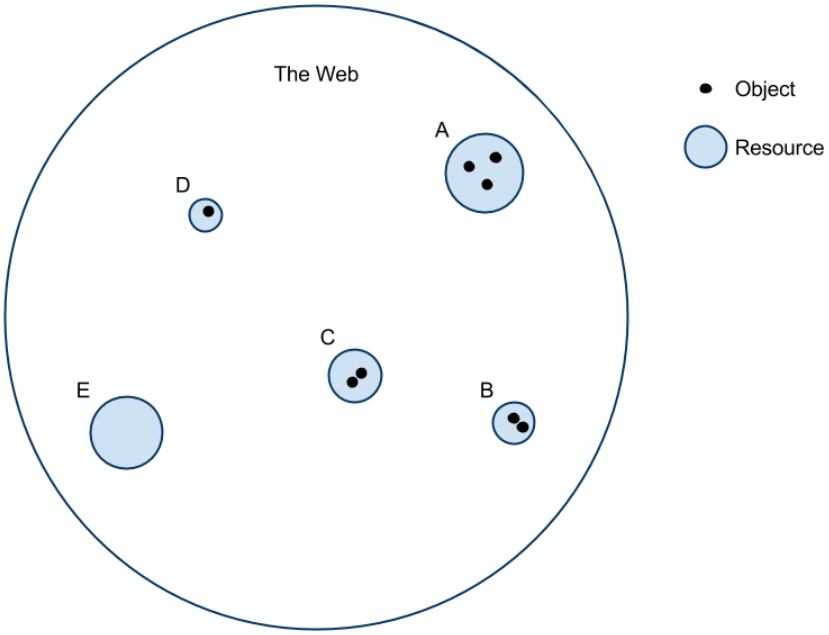


Figure 3.4. Resources and objects from a web perspective.

Figure 3.4 illustrates five resources on the web from a web perspective, where A and E are large resources and B-D smaller resources. The dots in A-D illustrates that objects in the resources are possible to reach directly from the transparent web. Resource E has no dots and that means the content in E only can be reached from within E (the deep web). The content in E may be stored in a database and demands a query to be displayed. This perspective highlights how much the user has to choose from, but at the same time some objects need searching in several steps to be reached. Different search strategies are needed to reach the objects available on and via the web (Fransson, 2011). The findability of the objects is supposedly higher in A-D than in E, which is not findable directly from the web.

The web is made up of enormous amounts of information published on web servers. How large the web is not known. The size has been estimated in different ways, in number of pages or in terabytes. A recent estimation is 600 billion indexable pages (Levene, 2010). In 2008 Google had encountered one trillion unique URL's (Alpert & Hajaj, 2008). On a higher level the objects and resources can be related to the Bow Tie model of the Web (Broder et al., 2000), primarily in the right part due to the linking to and between the objects. The model of the web is based on a crawl of the web done by the search engine Alta Vista done in 1999. The central core is made up of pages linking to each other, the pages has both inlinks and outlinks, and forms the weave of the Web. In the left part of the model there are pages which links to pages in the central core, but has no inlinks (page A in Figure 3.1). To the right in the model are the pages which has inlinks but no outlinks (page C in Figure 3.1). There is also other linking to pages not covered by the crawl,

and some pages are disconnected, with neither inlinks nor outlinks (page I in Figure 3.1). The characteristics will be discussed later in this chapter in relation to structure and content in general and the studied cultural heritage resources in specific.

3.2.3 Related web concepts

There are several concepts that are related to web findability. The concepts can be divided into three categories: technology focused, user focused, and mixed focused. Two similar technology focused concepts are *the deep web* and *the invisible web* (discussed in the previous section). The concept *web visibility* is used by for how well a page or object ranks in web search engines. In this sense web visibility is the outcome of the process of SEO – or lack of SEO. Web visibility focuses on visibility on the web, in web search engines. A similar concept is *digital visibility*, but for the visibility within resources. Digital visibility is used by the CIBER research group and focuses on how visible objects are within a web site (Huntington et al., 2004). An object which is promoted with a short text and a link at the top page of a site has a higher degree of digital visibility than objects that kind of promotion.

A user focused concept is *cognitive invisibility* (Ford & Mansourian, 2006). Cognitive invisibility is related to information sources a user did not find, either missed or unknown/unavailable depending on the user's level of uncertainty that the information relevant to their queries was on the web. Cognitive invisibility is close to two other concepts: *Personal space of information*, a concept in *Personal Information Management, PIM* (Jones, 2007, 2008); and the close couple *Information source horizon* (Savolainen, 2008) and *Information horizon* (Sonnenwald & Wildemuth, 2001). The later concepts focus on where the user actually looks or searches for information, contrary to cognitive invisibility (see Chapter 2). Neither of these concepts are system measurements, and they are instead related to the user's background knowledge and level of search skills.

A new concept is the *Filter bubble* (Pariser, 2011), which is an individual web space. The individual web space is created by technology in interaction with user behaviour, a process of automatic personalisation. Web search engines for example adapt the results in the search engine results page based on the user's settings, previous searches and location, so every searcher get a customized result.

Findability is also related to *accessibility*. Web accessibility refers to how accessible a website is for people with disabilities, e.g. a visual impaired person using text-to-speech software, but it is a technology focused concept. Around web accessibility there are guideline, standards and in some countries legislation (Thatcher et al., 2006). Kopackova, Michalek and Cejna studied the accessibility of local e-government websites and included findability as a part of accessibility. They argued that it takes information literacy to find something on the Internet, and thereby limited information skills could be considered as a disability and then findability requirements is

a part of accessibility. In the study they defined nine criteria for accessibility and findability testing (Kopackova et al., 2010) as shown in Table 3.1.

Table 3.1. Accessibility and findability in a local e-government study (Kopackova et al., 2010), the division between accessibility and findability is added by me.

No	Criteria	Accessibility	Findability
1	Findability through Czech search engines		X
2	Page rank		X
3	URL comprehensibility	X	X
4	Observance of WCAG rules	X	(X)
5	Comprehensibility of screen reader output	X	
6	XHTML validation	X	X
7	Display in low-resolution browser	X	
8	Separation of content from graphic part	X	
9	Metadata usage	X	X

As shown in the list above several criteria are important to both accessibility and findability. The conclusions were that the compliance to the W3C recommendations, WCAG rules and XHTML validation, was low. And the general findability in the search engines was surprisingly low for the studied web pages (Kopackova et al., 2010).

3.3 Measuring Findability

Findability is a discrete measure based on six criteria of the information resource: object attributes accessibility, internal navigation, internal search, reachability and web prestige. All six key concepts are based on a systems perspective. Each of them are measured by the aspect listed in Table 3.2, Subject Access Points (SAPs) are discussed and defined in Section 3.3.1.

Table 3.2. Measured aspect of each findability criteria.

Criteria	Measured aspect
Object attributes	Number of Subject Access Points (SAP), and if fulltext
Accessibility	Number of WCAG-errors
Internal navigation	Reachable with the internal navigation
Internal search	Reachable through internal search engine
Reachability	Possible to link to object
Web prestige	PageRank-value

The six concepts have been identified in the research and professional literature (Björneborn, 2004; Ding & Lin, 2010; Enge, 2009; Langville & Meyer, 2006; Levene, 2010; Morville & Rosenfeld, 2007; Thatcher et al., 2006; Walter, 2008; Witten et al., 2006; Wormell, 1985), which all form part of the web findability. Below they are grouped on three levels: on-page, site structure and web presence, levels corresponding to the layers in the object model (Figure 2.7a). Often information seeking and searching are studied without taking the information system into account, especially on the web where the web search engines plays an important part. This is an attempt to

study the information searching behaviour in the context of findability, and thereby take some of the complexity on the web into account. In order to measure findability, and especially web prestige, the structural aspects are seen as a closed world, which is the webometric assumption necessary to calculate PageRank-values.

The degree of findability is a crucial aspect of information being found on the web. Different aspects of findability are important in information seeking on the web respectively within the resource. The total findability of a web page or file is a combination between *internal findability* within the website and the *external findability* which depends on the degree of integration between the website and the web (including web search engines), the presence of the resource on the web. Internal findability largely depends on the information architecture of the site, but also on the possibilities and limitations of the content management system, which can be seen in the navigational possibilities. External findability depends on the site's impact (prestige) on the web and the degree of search engine optimization. In both cases the accessibility and content of the objects are crucial because in combination they creates access points to the objects.

A simple example of the external findability is if a picture has an extensive amount of metadata, but it is not indexed by the large web search engines, the external findability would be considered low because of it would demand a direct link to the object to discover it. If the picture is indexed by the web search engines it would have high external findability due to the extensive metadata.

The six aspects (critera) are placed in different parts of the resource and object models (see Figure 3.5). After the presentation of the aspects they are grouped into on-page (objects attributes), site structure (internal navigation and internal search), and web presence (reachability and web prestige).

3.3.1 Object attributes

The on-page aspects concern the single objects, how the pages are viewed in the browser and what they contain. The aspects may depend on global settings within the resource, but they might also be completely object specific.

Object attributes is the content, form and format of the information objects. Important parts are metadata, file format and eventual full text. Text objects have different conditions than multimedia objects. The representations of the objects in the outer layer, in the systems hierarchy of links and internal search engine, and on the web in the form of anchor texts in inlinks and the description in Google's result list, are all built on the attributes of the object (Enge, 2009; Walter, 2008).

> "Theoretically speaking, the more access points [a system] provides, the easier it is for the user to locate the information in the system." (Chu, 2010, p. 221)

Documents are described by two different kinds of bibliographic languages; a formal language to describe the manifestation of the work and a subject or documentary language to describe the

content of a work. The vocabulary of a bibliographic language is called metadata and it is determined by metadata standards and other rule systems. The formal language describes the formal aspects like author, title and size of an object (Petras, 2006).

> "Documentary languages vary in vocabulary restrictiveness (from strongly controlled to not controlled terms), size (from a few terms per document representation to full-text) and language type (alphabetic words or classification codes) but also in syntactic, semantic and pragmatic rules (how they are structured in a document representation, how vocabulary terms are related to each other and how they are selected for document representation)." (Petras, 2006, pp. 64-65)

In the extreme end of documentary languages spectrum is full text, which is a non-controlled natural-language-based documentary language. "Using the full text of a document for its subject representation requires no effort for description; it provides, however, a rather unfocused view of the document's content" (Petras, 2006, p. 68).

Subject Access Points (SAP) are the metadata in form of different documentary languages for a single record, document or object (Wormell, 1985). In an objective way can SAPs be broad or narrow:

> "Different kinds of SAPs describe the subject of a given document in different ways, such as more or less exhaustive, more or less general or specific, in a more-or-less open or closed way, and so on." (Hjørland & Nielsen, 2001, p. 254)

The quote below is in theory obvious; the useful SAPs are more valuable than other SAPs:

> "The most valuable SAPs are those that make it possible for the user to identify the most highly relevant documents, that is, make the highly relevant documents the most visible in the database at the expense of less relevant documents." (Hjørland & Nielsen, 2001, p. 254)

In practice, when the SAPs are judged with subjective judgements the user, or the imagined user, is incorporated in the concept. The distinction between system and user is then repealed – which is natural in IIR – but the concept becomes blurred (like findability as discussed in Chapter 2). The SAPs, the description of the object in the documentary language, together with the formal, bibliographic language forms an *Access Target Area* (ATA). The ATA is a combination of subject access points and structural access points. Examples of structural access points are the title of the journal an article is published in or the web domain the web page belongs to. The more access points in an ATA makes the object easier to find compared with an object with a smaller ATA (fewer access points). The ATA is the quantitative measure of the access points. The relevance of the access points for the user is determined by the information need and is placed in the upper half of Figure 2.11 (the O marked part), while the ATA is defined as a query independent aspect and a part of findability (in the lower, X marked part of Figure 2.11).

Linguistic SEO is the practice of optimizing the use of keywords in text and metadata. Both using the same vocabulary as the perceived target groups and using keywords in a relevant way for both

human users and search engine crawlers (Nielsen & Loranger, 2006). Creating content or writing copy is not a relevant aspect of findability in this study because the content is fixed; it is the digitized cultural heritage. Writing descriptions of pictures or transcriptions of audio or video files is not a possibility due to the enormous amounts of digitalized material. Even the attached metadata must in some cases be generated automatically.

The most important aspect of the object is the number of SAPs, Subject Access Points, like descriptors/subject headings and other metadata. As described above the documentary language can be of different kind, from classification codes to full text. In the case of full text the text itself is the SAPs. The drawback of full text-SAPs are their level of abstraction, the topic is not described on a more general level and is not found with abstract or general queries. On the web there are three types of or places with "on-page" SAP: on the page, the title, and places with the metadata fields in the header of the html-page. There are also "off-page" SAPs on the web. Anchor text within hyperlinks on other pages pointing to an object might be seen as SAPs, and are used by the major web search engine in their relevance ranking, and the objects post in i.e. Europeana, Flickr, etc. Full text is in the thesis seen as a special kind of SAP and is counted as an addition to the number of SAPs. Each studied object is evaluated according to the two findability measures:

- How many SAPs do the object contain?
- Is the object in full text?

How the measuring is done is discussed in Section 5.3.2.

3.3.2 Accessibility

Accessibility concerns the possibility to access information objects. The research concering accessibility focuses on accessibility issues for persons with disabilities, but a disability-friendly website is also a search engine friendly site. The search engines are limited in their way of interpret the content on web pages, just as a speech synthesizer that reads the web page content. Techniques like flash and java script may limit or totally prevent the access to the information objects. Accessibility is closely related to usability (Thatcher et al., 2006; Walter, 2008).

Good accessibility for disabled users is positive for both normal users and search engine spiders. Descriptive keywords are one example of an aspect of accessibility standards that will help all kinds of users to better understand the contents. Generally all web standards improves the findability of a page as they brings order and hierarchy to the information (Walter, 2008).

The Web Content Accessibility Guidelines 2.0 (WCAG 2.0) is a set of guidelines for making the content on websites accessible to people with disabilities. The guidelines based on the four principles of accessibility (W3C):

1. "Perceivable - Information and user interface components must be presentable to users in ways they can perceive."
2. "Operable - User interface components and navigation must be operable."

3. "Understandable - Information and the operation of user interface must be understandable."
4. "Robust - Content must be robust enough that it can be interpreted reliably by a wide variety of user agents, including assistive technologies."

It is debated how measureable the guidelines are, WCAG 2.0 contains subjective, rather than objective, measurements (Thatcher et al., 2006). The web browsers compensates for bad html code, but still "most Web sites have accessibility barriers that make it either difficult or near impossible for many people with disabilities to use these sites" (Harper & Yesilada, 2008, p. 62). In the present study is the following accessibility measure used:

- How many WCAG-errors does the object contain?

How the measuring is done is discussed in Section 5.3.2.

3.3.3 Internal navigation

The site structure aspects concers features valid for whole levels of a resource or for a type of objects.

Internal navigation is a central part of an information system on the web. It is about link structures and the possibility to follow links down to individual objects. On websites navigation by links is the main way of navigation (Ding & Lin, 2010; Morville & Rosenfeld, 2007).

Having a link structure to websites for search engine crawlers to follow is the basis for internal link navigation. The links should be in plain HTML and look like they describe static web pages. Describing target pages in anchor text internal links. Link structure should reflect hierarchy of website. The navigation aid bread crumbs on top of each page is one solution, it provides links to all pages higher up in the hierarchy (Nielsen & Loranger, 2006).

When it comes to the acceptable number of steps or clicks from the top page there are different schools of thought. Some meant earlier that hierarchies should be broad instead of deep; all content should be easy to reach for the user (Straub & Weinschenk, 2003). Another point of view is that it does not matter how many steps the user is forced to go to reach the content as long as the information scent is increasing, the user experience himself coming closer to the desired information (Spool et al., 2004). In the present study I have chosen to just measure if an object is possible to reach with the internal navigation:

- Is it possible to reach the object with the internal navigation?

3.3.4 Internal search

Internal search is the second way of navigation within websites. How well an internal search works depends on the search engine's performance and settings together with the attributes of the objects (Ding & Lin, 2010; Morville & Rosenfeld, 2007). To study the performance of the internal

search engines is a large task; here I just study if the objects are possible to find with the internal search.

- Is it possible to find the object through the internal search engine?

The site structure aspects are based in two areas: information architecture for the organization of content and the internal navigation, and IT for the implementation of internal search and general web server settings. For a human user it is required that an object is findable either through the internal navigation or by internal search, otherwise the findability score is zero regardless of the other aspects.

3.3.5 Reachability

Web presence concerns the findability aspects outside the resource and if the object are possible to reach from the web, not from within the resource. *Reachability* is a requirement for both users and search engines to reach the objects. If no links to the object exists it will be an isolated island on the web. Under the concept of reachability falls the indexability by search engines, so the objects can be found in the search engines (Björneborn, 2004; Enge, 2009; Levene, 2010; Walter, 2008). The requirement is that the URL has to be stable and unique, so it is possible to link to the object. If the URL is static or dynamic is not important on the Web anymore, it was a crucial factor earlier when the web search engines were less sophisticated. Web sites using frames for displaying several web pages at once in different parts of the window is a common cause of low visibility. The main page containing the frames is reachable , but the pages in the frames of the main page has no public URLs and are not possible to link to in an easy manner. I have chosen to examine the URLs and the possibility to link the objects as the reachability-measurement:

- Is it possible to link directly to the object?

3.3.6 Web prestige

Web prestige is the measure of how many and who which links to an object on the web. The prestige of the inlinks is an important factor beside the number of inlinks. A prestigious inlink give more reputation than many links with low prestige. Through link analysis it is possible to measure the prestige of an object. The measurement can be seen as the possibility to land at the object at a random choice. The most widely known measurement of web prestige is PageRank, which is a part of the relevance calculation in Google web search (Enge, 2009; Langville & Meyer, 2006; Levene, 2010; Witten et al., 2006). I have chosen to use the PageRank-value of the objects as an approximate value of the web prestige:

- What it the PageRank-value of the object?

How the measuring is done is discussed in Section 5.3.2.

3.3.7 The total findability indicator

The six aspects together give an indication of how easy it is to find a specific object on the web. Findability is the sum of the six aspects and forms a measurement that is possible to relate to both the use of information objects and resources, and the search strategies chosen by the users. It is also possible to compare the findability of similar objects in different information systems, e.g. pictures of a church published both in the social sharing service Flickr and in a dedicated resource with digitized cultural heritage. Findability, as it is defined here, gives an indication of how likely it is that a user searching for images of a church can find a specific image in relation to similar images on other sites. The six aspects of findability are playing important roles on different levels, and the levels generally correspond to the layers in the object model as shown in Figure 3.5.

The six aspects of the findability can be connected to the navigation strategies. They are also related to information search skills, such as to the four central sub competencies in information seeking on the Web that was described in Chapter 2 (Gerjets & Hellenthal-Schorr, 2008). Every aspect of the findability can be viewed as a specification of each sub competence. For example, the first of the sub competence, *Media background knowledge*, can be divided into knowledge about web prestige, reachability, and so on.

In Figure 3.5, the six findability concepts are positioned relative to object, information resource and the web. Together with Figure 4.3 the model represents two layers that illustrate the findability concepts that are essential for each navigation strategy (and vice versa). For example, the accessibility, the reachability and Web prestige are important in navigation via links. But for navigation by using a search engine the characteristics of the object are of great importance when the search engine relevancy is estimated by matching keywords with the content and metadata of the object.

In Figure 3.5 the findability aspects are placed in the object model and the resource model to highlight their individual importance for the findability. Some aspects are placed on lines in the model to show that the aspect has a bridging role between layers (in the object model to the right), right between the levels of the resource (C and D in the resource model) or at the border between the resource and the web (E in the resource model).

The findability aspects (as named in Figure 3.5):

A. Object attributes
B. Accessibility
C. Internal navigation
D. Internal search
E. Reachability
F. Web prestige

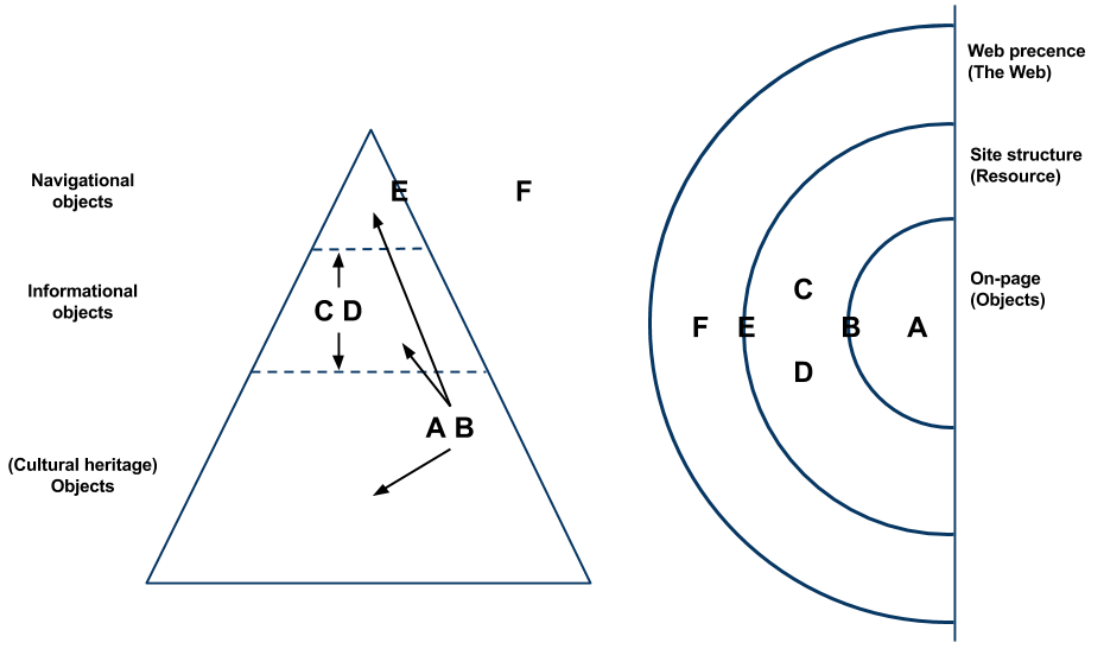


Figure 3.5. The aspects of findability places in the triangular resource model (left) and the object model (right) (Figure 2.7a+b).

The aspects of findability can be combined into different measures of findability, for example *internal findability* and *external findability*. Internal findability is the sum of the relevant aspects when the user is already in the resource; then the web presence is unimportant. Internal findability is findability within a resource or site, the sum of aspect A-D. *External findability* is the sum of the aspects important for finding the object from the web. The internal navigation and search are not important when an object is reachable directly from the web when the user in not in the resource. External findability is how findable the objects are from the web without taking the internal navigation possibilities into account: the sum of A, B, E and F. *Total findability* is the sum of all six findability measures, in the text just called *findability*.

3.4 Cultural heritage resources

3.4.1 Definitions of cultural heritage and digital heritage

Cultural heritage is a concept that is often used, but it is not defined very often. *The Danish Agency for Culture* (or the former *Heritage Agency of Denmark*) does not define the concept of cultural heritage (Jensen, 2008). Nor is CH defined in a Danish anthology about dissemination of cultural heritage (Lund et al., 2009). Hyvönen has recently defined CH:

> "*Cultural Heritage* (CH) refers to the legacy of physical objects, environment, traditions, and knowledge of a society that are inherited from the past, maintained and developed

> further in the present, and preserved (conserved) for the benefit of future generations." (Hyvönen, 2012, p. 1)

CH can be divided into three subareas according to Hyvönen (2012), which is a refinement of the UNSECO categories of CH (Unesco, 2008):

> "1. **Tangible cultural heritage** consists of concrete cultural objects, such as artifacts, works of art, buildings, and books.
>
> 2**. Intangible cultural heritage** includes phenomena such as traditions, language, handicraft skills, folklore, and knowledge.
>
> 3. **Natural cultural heritage** consists of culturally significant landscapes, biodiversity, and geodiversity." (Hyvönen, 2012, p. 1)

As seen in the categorization above cultural heritage is used on different levels, from an overall concept to cover something worth preserving to intangible in form of rituals and oral traditions. A definition from UNESCO is:

> "The cultural heritage may be defined as the entire corpus of material signs - either artistic or symbolic - handed on by the past to each culture and, therefore, to the whole of humankind. As a constituent part of the affirmation and enrichment of cultural identities, as a legacy belonging to all humankind, the cultural heritage gives each particular place its recognizable features and is the storehouse of human experience." (Unesco, 1989, p. 57)

The digital parts of the cultural heritage form another concept. UNESCO uses the term *digital heritage* and defines it as follows*:*

> "The digital heritage consists of unique resources of human knowledge and expression. It embraces cultural, educational, scientific and administrative resources, as well as technical, legal, medical and other kinds of information created digitally, or converted into digital form from existing analogue resources. Where resources are "born digital", there is no other format but the digital object.
>
> Digital materials include texts, databases, still and moving images, audio, graphics, software and web pages, among a wide and growing range of formats. They are frequently ephemeral, and require purposeful production, maintenance and management to be retained.
>
> Many of these resources have lasting value and significance, and therefore constitute a heritage that should be protected and preserved for current and future generations. This ever-growing heritage may exist in any language, in any part of the world, and in any area of human knowledge or expression." (Unesco, 2003, Article 1 - Scope)

"Digitization" is more often used instead of "converted", which is used in the definition above. In the process of digitization a digital copy of the original, physical object or artefact is produced.

These digital substitutes are shall not be mixed up with objects that are born digital and have never been physical. There are three types of digital heritage, digitalized and born digital, as mentioned in the UNESCO definition above, but also digital information about cultural heritage (both digital and physical). This heritage records or metadata is crucial in the information search process to find the other types of digital or physical heritage.

The heritage in digital form is very heterogeneous. The Danish board working with the digitalization of cultural heritage has the following categorizations of the heritage that can be digitalized (Kulturministeriet, 2009):

1. Printed texts, archive records and manuscripts
2. Photos
3. Music and sound
4. Moving images

The categorization above lacks information about the physical heritage: the immovable, the underwater, the intangible, and the natural cultural heritage. The categorization is not intended to contain that category of material, but if it is complemented with a fifth category, metainformation, will the categorization cover information on all types of digital heritage:

5. Metainformation

The division is important both in terms of user's information behaviour and how the present thesis is planned because the different categories are also linked to different research areas, i.e. image information seeking or multimedia retrieval. Many resources from the ALM sector contain just metainformation, for example online library catalogues.

3.4.2 Digitized cultural heritage on the web

The cultural heritage institutions' relation to the web is a large and complex matter which is largely outside this project. Of the digital strategies for the cultural heritage among the institutions within the ALM sector is perhaps the absence on the Web the most common, probably out of fear for the material to be manipulated and used in inappropriate contexts (Dahlgren & Snickars, 2009). But cultural institutions are to adapt to conditions on the Web, it is the political will:

> "The first premise that users can use the digital heritage is that it is accessible via the Internet. [...] The accessibility may be that the user through a search for example in Google may find some information about the material with a link to the digitized image, audio, video or similar." (My translation from Danish) (Kulturministeriet, 2009, p. 49)

How accessible a resource is on the web is an important part of how found and used it and its contents are. Especially important are the general search engines on the web, as Mark Levene has described:

> ”Search engines are currently the primary information gatekeepers of the web, holding the precious key needed to unlock the web both for users who are seeking information and for authors of web pages wishing to make their voice heard.” (Levene, 2010, p. 64)

How CH resources are found and used is a relatively unexplored area. Previously, when heritage was not digitized by the same extent as today, scientists and conservation was the explicit targets of the efforts. Now the public is an important audience for the heritage collections and the dissemination need to find new ways to supplement the old. Perhaps this is not the heritage resources that are available digitally today that users are most interested in. The Danish broadcasting corporation (Danmarks Radio, DR) did in 2007 a survey about the cultural habits of the Danish people and their interest in cultural heritage project danskkulturav.dk.

> “About 80% of all survey respondents would either definitely or probably use the access to DR's archives. Only 8% could not imagine using DR's radio and television archives.
>
> [The offerings] that tastes mostly of popular culture, film, television, radio and music are also those which most Danes in the study could imagine using. About 70% could imagine using the offer of access to Danish music and approx. 65% could imagine using the offer of access to Danish film.” (My translation) (Wieland et al., n.d., "Kulturarvsprojektets muligheder")

The survey of DR shows that it is the popular cultural audio-visual heritage that the Danes are most interested in. Here there may be one conflict which inhibits the use of the material. Snickars believes that there is a “medial hierarchy within the archives sector” (Snickars, 2005, p. 215).

> “In conservation terms, there exists an unspoken distinction between high and low media, both in content and form, which depending on the classification ranks differently important as cultural heritage. Older media attracts more appreciation, but also a new film on film base enjoys for example greater cultural legitimacy than videotapes. Radio is considered more important than television, and contemporary art photographs are of course far more priority than mass-produced imagery.” (My translation from Swedish) (Snickars, 2005, p. 215)

The question is what users are looking for, and what they find when it comes to digital cultural heritage resources. How users search for images, audio and video with respect to the use of keywords have been studied in the web search engines (Spink & Jansen, 2004), but how users formulate their information needs as search queries are also an important issue in the use of cultural heritage.

How they want to use the cultural heritage they find is a different matter. Today, the user can no longer be considered only as a user who passively consume what by the producer (the cultural institution) has produced, but is increasingly being simultaneously producer. This double or new role is sometimes called *produser* (Bruns, 2008).

In the Scandinavian countries the word *förmedling* (in Swedish) or *formidling* (in Danish) is used as a broad concept for spreading, mediate, disseminate and make culture available. There is no

exact translation into English. Holdgaard and Simonsen discusses formidling and the possible equivalents in English; dissemination and communication (Holdgaard & Simonsen, 2011).

> "Etymologically, *formidle* in Danish means to act as a link or a connection between two parts. *Formidling* is an ambiguous term covering a broad spectrum of concepts from knowledge, education and learning to communication and is used within several scientific traditions (Gudiksen, 2005). *Formidling* can be understood as one-way transmissions, as reciprocal exchanges and interpretations of meaning and as (inter)actions and change." (Holdgaard & Simonsen, 2011, p. 103)

In the article Holdgaard and Simonsen chose to use *formidling* together with communication instead of communication and dissemination. In the present text *mediation* is used to for the Scandinavian concept of *formidling*. In the thesis the mediation strategies for CH on the web are divided into two dimensions. The strategies can be either active or passive, and either focused or non-focused (general) (Table 3.3). Active resources can be active in two ways, both on the institutional side (based on a story line) and on the user side (user generated content). Passive resources are generally not continuous developed after they are published. Focused are resources with a narrow topic or a well-defined target audience. Non-focused are resources with a more general topic or several topics for a broad audience. In a museum practice two opposite extremes on how the museums utilizes the web are either *the internet as a channel for mediation* or *the internet as a possible space for exhibitions* (Løssing, 2008). The first of the museum practices is overall a passive and general approach to mediation, whereas the second practice is more focused and active.

Table 3.3. Examples of online Danish cultural heritage resource with different dissemination strategies.

	Active	**Passive**
Focused	Danmark set fra luften[5]	Forsvarets biblioteks digitale fotoarkiv[6] Guaman Poma[7]
Non-focused (general)	1001 fortaellinger[8]	Arkiv for Dansk Litteratur (ADL)[9] Kunstindex Danmark (KID)[10]

Another important aspect for the use is the copyright on the digitalized heritage objects. Contents with generous creative commons licenses might be used more because then it is possible to

[5] http://www.kb.dk/danmarksetfraluften/

[6] http://www.foto.fak.dk/fotoweb/

[7] http://www.kb.dk/permalink/2006/poma/info/en/frontpage.htm

[8] http://www.kulturarv.dk/1001fortaellinger/

[9] http://adl.dk

[10] https://www.kulturarv.dk/kid/Forside.do

publish remixed content in a legal way. The goal of the mediation might differ. The explicit political goals with the digitalization of the cultural heritage in Denmark have been a question of Danish identity, to marketing Denmark, and to fuel the creative economy. The internal goals in the heritage institutions might be on a more concrete level; to relieve the usage pressure on fragile physical object, or to supply the research with material that is easy to access.

3.4.3 Operationalization

The concept of cultural heritage is difficult to define and used in various ways. In many cases avoids the need to define what heritage is, in both policy and research (Jensen, 2008), as discussed in 3.4.1. A pragmatic definition is used in the thesis and departs from the collections of the so-called "cultural heritage institutions", i.e. archives, libraries, museums and institutions with audio-visual collections, and their collections are seen as cultural heritage resources. Furthermore, only the digital portions of these institutions collections are included in the present study, the digital cultural heritage resources and the objects they contain. This operationalized definition focuses on the "official" or state funded digitalised resources because it is of public interest to study the outcome of these digitalisation projects executed by governmental institutions, i.e. The Royal Library. Another reason is that the resources from the cultural heritage institutions are relatively easy to identify and can be expected to have at least some users, where small local initiatives might have relative few users. Without the anchoring in public institutions it is hard to draw a line between different kinds of resources on the web, and increasingly even cultural heritage institutions uses commercial platforms for mediation and distribution. If the platform or site hosting the cultural heritage resource is commercial the usage data will be real hard to get access to.

The study focuses on cultural heritage resources (according to the operational definition above), but on the resources level there are always alternatives to the resources from the cultural heritage institutions. An example is the various new web service that has emerged in the Web 2.0, such as Flickr[11] for pictures, YouTube[12] for video (and to some extent, sound), but also alternatives like Wikimedia Commons[13] (images). In some cases these new services are used by cultural heritage institutions to distribute cultural heritage objects, i.e. The Library of Congress makes photos available on Flickr Commons[14].

[11] http://www.flickr.com/

[12] http://www.youtube.com

[13] http://commons.wikimedia.org

[14] www.flickr.com/photos/library_of_congress

As stated in Section 2.3 the resources are made of objects on different levels. In the case of digital heritage the resources might be websites, sub sites, and online exhibitions. The common denominator is that the heritage resources are built around digital heritage objects, the objects are the core of the service and not some kind of add-on or extra feature.

3.4.4 The studied cultural heritage resources

The Danish digitalized cultural heritage is catalogued in *Kulturperler*[15], an online national bibliographic service maintained by the Royal Library. It contains over 200 different heritage resources, from small collections to vast archives. The main criteria when choosing which Danish online heritage resources to study are:

1. Well known or famous content
2. A substantial amount of content
3. A potentially large group of users
4. Potentially different/heterogenic user groups (e.g. not just students or hobbyists)
5. Non-temporary resources
6. Availible for research

The six criteria lead to exclusion of a large portion of the Danish heritage resources. Many resources are either narrow in scope (e.g. a large photo collection of old warships) or contains a small amount of content (e.g. 27 audio recordings of local dialects). In the end it had to be possible to get access to the log files as well as a possibility to cooperate with ALM institution regarding the web survey, before the resource cold be considered to be included in the study. There was a discussion with Danmarks Radio (DR) about Bonanza[16], a Youtube-like video archive, but it was not possible to study the usage of the resource. Based on the six criteria I have chosen to study the following three resources: *Arkiv for Dansk Litteratur*; *Kunstindex Danmark*; and *Guaman Poma Inca Chronicle*.

Arkiv for Dansk Litteratur (ADL)[17], in English *Archive for Danish literature*, is a web site for classic Danish literature. It portraits 78 authorships from the twelfth century up to 1938 (due copyright reasons) and contains all their works in full text together with documentary and literary-historical aspects. It contains over 160,000 seachable pages of classic Danish literature. The selection has been made by the Royal Library and *Det Danske Sprog- og Litteraturselskab*. All

[15] www.kb.dk/da/materialer/kulturarv/

[16] http://www.dr.dk/Bonanza/index.htm

[17] http://adl.dk/

the author portraits are written by subject experts on the authorships. ADL is chosen as one of the studied resources because it full fills all four criteria, and it contains a large amount of full text.

Kunstindex Danmark (KID)[18], in English *Art Index Denmark*, is a database covering a large part of the holdings in the Danish Art Museums. Besides the holdings the website contains information about artwork and artists. The site, Kunstindex Danmark, is administrated by a governmental agency, Kulturstyrelsen. It contains 239,000 works of art, many with thumbnail pictures, together with information on the artists and museums. The website also includes a dictionary of Danish artists, Weilbachs kunstnerleksikon. The reason for studying KID is that the resource represents a large class of museum catalogues available online, as well as meeting all four criteria above.

Guaman Poma Inca Chronicle (Poma)[19] at the Royal Library is a widely used digitized book; maybe the most read online e-book in Denmark. As a single work the chronicle contains almost 1200 pages and 400 large illustrations. It is a unique manuscript written in Peru in the beginning of the seventeenth century by Felipe Guaman Poma de Ayala. The chronicle was a letter to the Spanish king about the hard situation for the natives in Peru. It is the only work in the world that illustrates the life in Peru before the Spanish conquest and how the colonial administration of the former Inca society worked. "Poma" is on the UNSECO memory of the world list. The resource has two interfaces, one in English and one in Spanish, as the audience is primarily international. The Chronicle is included as one of the studied resources because it represents a common type of digitalised resources, digitalised manuscripts and books. The content of the resources is less general than in the other two studied resources. The content is unique and only available on the web site of the Royal Library and is thereby interesting to study.

There are both similarities and differences between the three studied resources. All three contains a lot of *full text*. In ADL there are digitalized texts by the included authors, plus the descriptions of the authorships. Poma is similar, digitalized pages from the manuscript, plus comments on the content. In KID are there descriptions of the artists, the part that is called Weilbachs art lexicon. The full text is not dependent on metadata, if it is indexed and searched for in the corresponding language. There are also *pictures* in two of the resources. In KID there are pictures of art work (thumbnails) together with some metadata. In Poma are almost 400 of the 1200 pages full page illustrations. The pictures are dependent on metadata to be found, or to be found through the surrounding text. None of the resource are included in Europeana.

In terms of the categorization of types of digital heritage objects the study deals with three types: text (type 1); photos (type 2); and metainformation (type 5). Limiting the object types to three gives the study both depth and complexity, but it is still manageable. If all the five types were

[18] www.kulturarv.dk/kid/Forside.do

[19] www.kb.dk/permalink/2006/poma/info/en/frontpage.htm

studied it would be hard to handle and compare the different findings. In Appendices 1-3 there are screen dumps of the objects in each resource studied in the findability analysis.

All the resources are passive, they are stable and most of the content is fixated (Table 3.3). There have not been any changes or re-designs since the launch in neither one of them. In KID artworks are added as the collections of the museums changes. It was on beforehand known that all three CH resources were used frequently and thereby possible to study the usage of CH objects in.

3.5 Chapter summary

Findability is a central concept when studying user's search behaviour on the web as it is a concept possible to apply both internally in a resource and on the web as a whole. The sub-research question focused on in the chapter was: *What aspects are important for measuring the findability of a web object (RQ1a)?* From previous research combined with literature from the professional fields of web design and search engine optimization six aspects of web findability were identified as central: attributes of the object; accessibility; internal navigation; internal search; reachability; and web prestige. The first two, object attributes and accessibility, are characteristics of the single object, whereas internal navigation and internal search are site or resource features which have implications on all objects within the resource. Reachability as also a structural feature of the resource, but it is one of two aspects concerning the web presence. The other aspect is web prestige, which is the only aspect that is based on external actors. The aspects important for finding objects within a resource are combined into internal findability. In the same way is external findability a combination of the aspects crucial for an object to be findable from outside the resource, from the web.

The second half of the chapter addressed cultural heritage and which resources that should be chosen among the available Danish resources online. Three different resources were included in the study based on four criteria: Well known or famous content; a substantial amount of content; A potentially large group of users; and, Non-temporary resources. The resources are Arkiv for Dansk Litteratur (ADL), Kunstindex Danmark (KID), and Guaman Poma Inca Chronicle (Poma). They are in different ways typical, and together they represent the Danish cultural heritage resources online in the present study.

4 Users in action

> "What the study of surfing reveals is not only a law that describe the way we hop from link to link, but also an interesting insight into human behaviour and the existence of a kind of economy of attention that guides our surfing." (Huberman, 2001, p. 42)

This chapter gives the theoretical background for the empirical research on the second research question, the question concerning how users find and use the resources. The main goal of the chapter is to identify indicators for research questions RQ2a-f, i.e. the user-side of the interactions in the URI model (Figure 2.11). The following question is answered in the end of the chapter: *What measurements or indicators of usage are central to measure the navigation to the resources and the actual use?*

There are three basic categories of measurement that can be derived from the conceptual framework in Chapter 2 when it comes to usage. As shown in the IS&R model (Figure 2.8) arrow number 2 illustrates the interactions with the information system. As shown in Figure 2.9 the interactions either can be with the web, with a resource or with an object. In fact every interaction on the web is with a local information system, but the distinction here is between the studied cultural heritage resources and all the other local systems (here called the web). In the IS&R model the central entity is the cognitive actor, the user. And the with ELIS framework the user and her behaviour is placed the in larger context of everyday life. Based on these starting points, and in the light of the second research question, this chapter contains the following sections: User characteristics and activity contexts (Section 4.1.1), and information searching as an activity. In the first section the user and her contexts are explored with emphasis on needs and tasks (Section 4.1.2), together with affordance as an extension of search skills (Section 4.1.3). In the second section information search activities are in focus (Sections 4.2.1 and 4.2.2), especially navigation to and within sites (Sections 4.2.3 and 4.2.4), and it ends with a sub-section on web search strategy indicators.

The first sections in the chapter are related to the URI model (Figure 2.11) through the letters A-G. When studying the general use of web resources based on log data different contexts may be identified, e.g. geographic location or organizational affiliation based on IP number. It is not possible to talk about information behaviour or practices grounded in more specified social context due to the nature of the data. The data is sparse in terms of both context information and user actions over time. The focus is information interaction and the single sessions.

4.1 User characteristics and activity contexts

User activities can be seen in two types of contexts, internal and external. Examples of internal contexts are information need giving rise to a search task (Byström & Hansen, 2005) or a knowledge gap (Dervin, 1998), but also learning style (Ford et al., 2002) and level of search skills (Hölscher & Strube, 2000) (factors within in the cognitive actor in Figure 2.8). Organizational and social belongings are external contexts (the right part of Figure 2.8). ELIS can be seen in both internal and external contexts. Everyday life activites can be routine or non-routine (Figure 2.3), which are a form of internal context. Everyday life in itself forms an external context, with for example family duties and social relationships in the larger context of socio-economic class. When it comes to user actions, interactions, with the information system the system itself forms a context (the left part of Figure 2.8). The content in the system, how it is organized and presented, forms a context. The content together with the functions of the system form a complex context that affects how the user acts.

All contexts affect the activities, but some are more important than others. The URI model (Figure 2.11) stresses two dimensions: query dependent aspects and query independent aspects. During the information search process (E in Figure 2.11) the information need (F) and the information skills (H) together affects what actions the user takes and what affordances she detects and makes use of. In the model the user and her contexts (I) are the source of the need and the skills.

Contexts are complex and can be many things. Pharo connected different factors in the information search process together in the conceptual SST-method scheme (Pharo, 2002; Pharo & Järvelin, 2004). The contextual categories, in addition to the search process, in the SST-method scheme are: searcher, work task, search task, and, social and organizational environment. Within the search process in the SST method scheme there is a distinction between *search transitions* and *search situations*, where transitions are navigation to information resources/objects and situations are interactions with objects with in a resource.

One model of the users' context in information seeking is developed by Kofod-Petersen and Aamodt (2003). They studied information seeking via mobile devices, but the contexts are the same in a web environment. The user contexts:

1. Task context
2. Social context
3. Personal context
    a) Physiological context
    b) Mental context
4. Spatio-Temporal context
5. Environmental context

The list illustrates the range of contextual factor which might be investigated (Ruthven, 2011). Both Kofod-Petersen and Aamodt, and Ruthven view the user from the system side, from a design or IR evaluation view point. In addition to the five contexts, task, social, personal, physical and

environmental, it is possible to add a sixth, the information system. The system itself is also a context influencing the user and the actions taken, at least when studying user behaviour during system interactions.

The majority of the contexts listed above are external contexts, contexts outside of the user. The social context often includes both the organizational and cultural contexts; they might be seen as specifications of the social context (Ingwersen & Järvelin, 2005; Pharo, 2002). Factors in the users' social context are domains, goals, work task situations, strategies, preferences and interests (Ingwersen & Järvelin, 2005). The work task situation might be a structured task in a work situation (Byström & Järvelin, 1995), but it can also be a task in everyday life (Savolainen, 2008) or a task related to hobby activities (Hartel, 2003). Hobby activities are a form of serious leisure (Stebbins, 2007), work-like tasks outside work (see Section 2.2).

The spatio-temporal (physical) context is especially important in situations with mobile devices, but it is also important in normal web based computer interactions (e.g. language settings can be based on location). Environmental might also be important in some settings, private information or company secrets might be inappropriate in public places (Ruthven, 2011).

4.1.1 User characteristics and information source horizon

Within the user (I in Figure 2.11) a large number of factors has been studied in relation to information seeking, for example learning style (Heinström, 2003), cognitive style (Ford et al., 2002), search expertise and subject expertise/domain knowledge (Hölscher & Strube, 2000), as well as the "normal" demographics, background data, of the users are: gender, age, level of education, profession, and geographic location. Occupation has been the most common "structure" for investigating information seeking, and the second most common is "role". Demographics' grouping is another "structure" for studying seeking behaviour. The most common variable to break down the studied participants into groups with is age, especially children, teenagers and elderly (Case, 2007). In the 2000s the "Google generation" has been in focus of the information seeking research (Pors, 2005; Rowlands et al., 2008).

The five contexts presented in the previous section (Kofod-Petersen & Aamodt, 2003) lead to that each user has individual information source horizons (Savolainen, 2008; Savolainen & Kari, 2004) or information horizons (Sonnenwald & Wildemuth, 2001; Sonnenwald et al., 2001). Every user has her preferred information sources which she is returning to when information needs are identified. The sources in the horizon limits the user when searching for information, some sources are always overlooked or unknown to the user, which is a constantly present dilemma in information searching on the web; the user can not have a complete overview over the available information sources (as discussed in Chapter 2). As opposite to the model of humans as "the rational man" Herbert Simon called this bounded rationality. Humans satisfice, which is defined as a process "through which an individual decides when an alternative approach or solution is sufficient to meet the individuals' desired goals rather than pursue the perfect approach" (Simon, 1971, p. 71).

Related to the information source horizons is the concept of *cognitive invisibility*. Cognitive invisibility is a cognitive subjective view of "web visibility" (Ford & Mansourian, 2006). The concept is based on two dimensions; if relevant information is found or not, and the level of uncertainty that the information exists on the web. When doing exploratory search on topics were the goal is not on beforehand known objects the degree of uncertainty that the sought information exists and is available might determine the motivation of the user and the strategies used.

4.1.2 Information need, task and intention

Information needs (F in Figure 2.11) has been expressed in different ways in information science research, as knowledge gaps (e.g. Dervin, 1998) or anomalous state of knowledge, ASK (e.g. Belkin et al., 1982). Taylor proposed a scale from not-yet conscious information needs to the need formalized as a query to a search system (Taylor, 1968). Ingwersen and Järvelin classifies information needs with three dimensions: the intentionality or goal of the searcher; the kind of knowledge currently known by the searcher; and the quality of what is known (Ingwersen & Järvelin, 2005, p. 291). The categorization lead to eight different types of information needs in a search session. But should not the intention follow on the information need, e.g. because of a specific information need the users executes a search with the intention of finding information solving the need? Or is the intention in the first case related to the work task and in the second case with the intention of the actual search? Intention is used to connect the information need with the search strategy and general information behaviour. Work task has become an important unit of analysis. Byström and Hansen has developed a conceptual framework for tasks in information studies (Byström & Hansen, 2005). An information seeking task is seen as a sub-task of a work task, and an information searching task is a sub-task of an information seeking task. A large or complex work task can require several information seeking tasks, which generates several search tasks.

The following list contains different kinds of search tasks which highlight the complexity of IS&R using the metaphor of the needle and the haystack. A searcher might search for:

- a *known* needle in a *known* haystack;
- a *known* needle in an *unknown* haystack;
- an *unknown* needle in an *unknown* haystack;
- *any* needle in a haystack;
- the *sharpest* needle in a haystack;
- *most* of the sharpest needles in a haystack;
- *all* the needles in a haystack;
- affirmation of *no needles* in the haystack;
- things *like* needles in any haystack;
- let me know *whenever* a new needle shows up;
- *where* are the *haystacks?*; and

- needles, haystacks – *whatever*. (Koll, 2000, "Information Retrieval Backdrop")

All the listed search tasks have different intentions and illustrate the complexity of search and the underlying factors. Queries formulated by users in search engines can be analysed regarding their underlying intention. Broder has proposed a taxonomy of web search based on a couple of hundred queries in the search engine Altavista (Broder, 2002). He identified three categories of searches with different intentions behind them. An *informational* query is the "traditional" information need which is met by topical information. A *navigational* query is intended to take the user to a specific site or page, often known on beforehand. In *transactional* queries the users have the intention to do some actions on the goal web page, e.g. download software or book tickets for a movie at the cinema. In the latter two query types informational queries are commonly included as a starting point (Alfort, 2013).

Several researchers have developed the taxonomy further. Levinson and Rose expanded the information category into several sub categories, e.g. "directed informational" searches when the user wants to learn something particular about the topic and "locate" searches for finding out about real world services and products. Transactional searches were renamed "Resource" and was divided into: Download, Entertainment, Interact, and Obtain (Rose & Levinson, 2004). Jansen and colleagues did a similar hierarchical classification of user intent expressed by queries in web search engines where they also divided the navigational searches into two subcategories. Navigational searches are either navigation to transactional or navigation to informational depending on the intention of the user after the navigational step based on intention (Jansen et al., 2008).

Another way of looking at queries is the principle of polyrepresentation (Ingwersen, 1996; Larsen et al., 2006). Within all three categories in Broder's taxonomy different aspects of the representation of information on the web can be used to specify search queries. The principle of

polyrepresentation has primarily been related to bibliographic searches in traditional IR settings for system evaluation. In Figure 4.1 the categories of representation are illustrated.

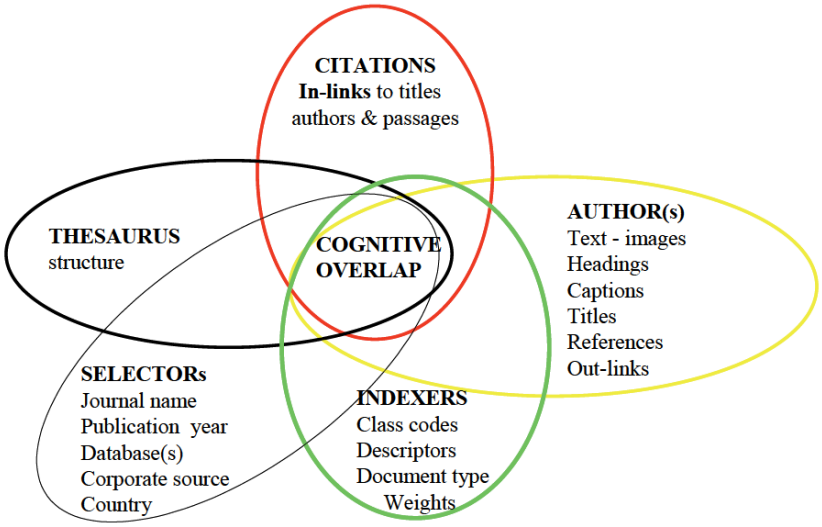


Figure 4.1. The principle of polyrepresentation of academic documents (Larsen et al., 2006, p. 89).

Searches with a web search engine works in the same way, e.g. the creator of the web content replaces the Author(s) in the model and the citations are replaced with hyperlinks. The controlled information architectural aspects, thesaurus and indexers, are missing on the web. The Selectors might correspond to the infrastructure of the Web with domains, subdomains, top level domains and country suffixes. The less structured representations of the information on the web can be used when searching, and often with success. An example of a search engine query is "Andersen site:adl.dk", which combines content (Andersen) with domain and top level domain (adl.dk as a limiter in the search engine) and just hits from adl.dk are presented in the results page.

The use of search terms was studied in the Getty project. In the project studied Marcia Bates with colleagues search behaviour of scholars in the humanities. In the first report is the search terminology of the scholars analysed, both the single search terms and the combination of search terms used in bibliographic databases (Bates et al., 1993). I will adopt a similar approach and classify the queries in the referring search engines. The queries will be classified on two levels. First the overall intention: informational, navigational or transactional (Broder, 2002). Secondly will the intentional queries, or the intentional parts of the queries, be classified according to the topic the search element in the query (in Table 4.1).

Table 4.1. Categories and subcategories used in the analysis of queries submitted in the referring web search engine.

Search goal (overall)	Short	Search element	Description	Example
Informational	I-CR	Creator	The creator used as search term(s). Author (ADL), Artist (KID) or Guaman Poma (Poma).	viggo kragh-hansen
Informational	I-SO	Specific object	A search for a specific object by title, e.g. an artwork or text.	Kukkenbjergvejenen aftenscene
Informational	I-FU	Full text	A quote	"han er mig kær"
Informational	I-GEO	Geographical	Name of a place, e.g. Region or Country	fyn
Informational	I-GEN	Genre	Different types of content, e.g. poem or self-portrait	komedie skagensmalere
Informational	I-TP	Time period	A period of time or a specific year. Might also be a literary period or a phase in history.	senklassicisme
Informational	I-SI	Specific institution	e.g. museum	skovgaardmuseet
Informational	I-TO	Topic	About the topic of the content.	forfatterskab
Informational	I-TY	Type of material	The kind of object search for, e.g. drawing or full text.	drawing foto
Informational	I-OT	Other keyword	Topical keyword not belonging to any other category.	baron painter realist
Navigational	N	-	The name of the resource or parts of its URL as search terms	adl weilbachs kunstnerleksikon
Transactional	T	-	The search terms indicated some kind of activity, e.g. download.	(no present in sample)

The classification of the queries will reveal parts of the information need and underlaying intention of the users.

4.1.3 Search skills and the web

When a user navigates the web a lot of skills and knowledge come into play. In Library and Information Science and among librarians is *information literacy* often used to cover these essential competencies for information searching, sharing and using (Chevillotte, 2010). Information literacy is thus a complex of knowledge, skills and understanding of information practised in different situations. It is discussed how generic or situation dependent information literacy is. There are many attempts to define what information literacy is or which competencies

an information literate person has, e.g. the definition from *The Association of College and Research Libraries* (ACRL). The majority of the definitions of information literacy based on a tradition of user education about information sources, techniques and evaluation of information, primarily in libraries at educational institutions (Limberg et al., 2009). The definitions can be said to be based on a structured paradigm in which information has traditionally taken place in controlled information systems, e.g. bibliographic databases or physical library, and search experts played a major role. The development of new information systems, particularly the web, has led to a new paradigm, the web paradigm, where information search is more iterative and is made by end users themselves (Pharo, 2008). Through its hypertext structure the Web enables also explorative search in greater degree than in more traditional information systems (White & Roth, 2009).

The focus is on the particular context the web form, and thus mainly the web paradigm. Gerjets and Hellenthal-Schorr have developed a method for training the students' information literacy skills on the web (Gerjets & Hellenthal-Schorr, 2008). As a starting point for their work, they have deconstructed different definitions of media literacy and information literacy, and divided and grouped the partial competences from the different definitions. Then they have put together the four main partial competences in information searching on the web:

> "Media background knowledge: Background knowledge with regard to the development and structure of the Internet and with regard to specific features of the WWW as information environment.
>
> Media operation skills: Skills for using computers, the Internet, and the WWW (e.g. how to connect to the Internet, how to use a browser software, how to use search engines and other web tools).
>
> Orientation skills: Ability to keep oriented with regard to the information sources provided by the WWW.
>
> Selection and evaluation skills: Ability to evaluate information provided in the WWW with regard to its relevance in the context of a current information problem as well as with regard to its quality and credibility. Ability to select information according to these evaluation criteria." (Gerjets & Hellenthal-Schorr, 2008, p. 696)

In the present thesis *web search skills* is used instead of information literacy or information retrieval competencies, and it is seen as a generic concept., which is partly transferable between different media or situations. Ingwersen and Järvelin suggests two types of domain and IS&R-knowledge, declarative and procedural knowledge. Declarative knowledge is the passive knowledge about the information system and how to perform searches in terms of declarative IS&R-knowledge, and declarative domain knowledge is about, or embedded in, information objects within the domain. In the same manner is procedural knowledge divided into IS&R- and domain knowledge, and the procedural knowledge is activity-related. An example of procedural IS&R-knowledge is search task execution skills, and problem or work task solving is examples

of procedural domain knowledge (Ingwersen & Järvelin, 2005, pp. 46-47). The division between domain knowledge and IS&R knowledge corresponds with query dependent and query independent aspects discussed in Section 2.5.1. The knowledge types are placed within the user in Figure 2.11. In other words declarative knowledge is how much a person knows about a subject, while procedural knowledge is the competency to perform tasks.

The generic nature of web search skills, a type of IS&R knowledge, is one of the greatest strengths of the concept, and without transferability between various media losses the concept some of its value. At the same time every situation where information skills are used unique and therefore are different combinations of knowledge and skills in the play every time. This means that the individual's actual level of web search skills varies depending on the situation.

In the information search process the user is interacting with the information space. The interactions and the information (infra)structures together form an information ecology (Huvila, 2009). The term affordance is used to describe user potentials. Dourish describes affordance as "a three-way relationship between the environment, the organism and an activity" (Dourish, 2004, p. 11). In the thesis the web and local information systems is the environment, the user is the organism and the activity is navigation and search on the web. The quote from Dourish above does not describe the nature of the affordance. Gibson, who introduced the concept of affordance, describes the nature of affordance as:

> "An important fact about the affordances of the environment is that they are in a sense objective, real, and physical, unlike values and meanings, which are often supposed to be subjective, phenomenal, and mental. But, actually, an affordance is neither an objective property nor a subjective property; or it is both if you like. An affordance cuts across the dichotomy of subjective-objective and helps us to understand its inadequacy. It is equally a fact of the environment and a fact of behavior. It is both physical and psychical, yet neither. An affordance points both ways, to the environment and to the observer."
> (Gibson, 1979, p. 129)

In Gibson's definition of affordance the users meets information systems (subjective and objective) with behaviour and environment (psychological and physical/digital). In the URI model (Figure 2.11) are these meetings seen as the potential overlap between the individual user's search skills and objects findability in the lower query-independent part of the model as affordances (the X marked structure level). Information search skills enable affordances of the web as an information system; information becomes accessible and usable, and determines how the retrieved information can be used. The findability of the objects on the web enable affordances in that the information objects becomes a part of the user's environment, they becomes possible to find.

In Figure 2.11 affordance works on two levels. On the query-dependent level affordance enables in the first phase the "discovering" and comprehension of the representation of the information objects. In the second phase is the representation in the query-dependent level (the O marked content level) evaluated for relevance by the user. Then, if the information is found relevant, a

cost-benefit analysis is done by the user to determine if the relevant information found is worth the effort. On the query-independent level search skills are determining the affordances perceived in the information system. The perceived affordances enable or limits the action possibilities "offered" in the given situation.

Affordance might be compared with the tension in an electromagnetic field; it binds the actions of the user with system, in both in Gibson's and Norman's notions of the concept. Norman has a designer approach to concept and sees affordance as properties of the objects (Norman, 1988). In later writings Norman focuses on "perceived affordances" instead of affordances in general, a term he means that the design community should use instead of just affordances (Norman, 1999). The system has certain properties and an intended use (Norman's view) but the process of interaction becomes larger than the properties of the system when the user comes into the picture, with her information need, search skills, motivation, etc., and the subject and object becomes integrated (Gibson's view). McGrenere and Ho developed an affordance framework for design based on the distinction between the degree of affordance and the degree of perceptual information (how easy it is to perceive the affordances) (McGrenere & Ho, 2000). Affordances occur during the information search process and are perceived by the user as action possibilities, the basis for interaction.

4.2 Information searching as an activity[20]

Information searching is activities which with the goal to find information of some kind, and the linked individual activities constitutes a search process. The search process is what happens in the encounter with the information system, the interaction between user and system. The user initiates the interaction with an intention, a need for information in a broad understanding. The need may be explicit and well defined, but it can also be very unclear. The user wants to know something, find answers, locate an object or extend their knowledge. To solve the information needs of the user initiates an information search (Ingwersen & Järvelin, 2005). Marchionini describes information literacy as a set of skills and concepts, while information searching is a fundamental human process. To develop students' understanding of the information retrieval process is therefore one of the key elements of the teaching of information literacy (Marchionini, 1999).

The information search process consists of eight sub-processes according to Marchionini, from the recognition and acceptance an information problem, to execute a search, extract information

[20] In the following section I refer to Marchionini and his use of the concept of *information literacy* to describe a set of skills and concepts. I have not replaced information literacy when discussing Marchionini's ideas to keep the clearity of Marchionini's original texts despite the use of *web search skills* in the thesis.

and stop (Marchionini, 1995). The sub-processes are not necessarily in a linear sequence, the transitions between sub-processes can vary. How the sub-processes are carried out depends on the personal information infrastructure that people develop, an infrastructure that consists of several parts (Marchionini, 1995):

- Mental models of knowledge domains, search engines and past information searches;
- General cognitive skills and information seeking specific skills;
- Attitudes and mental control mechanisms;
- Material resources such as time, money and equipment.

The first infrastructure refers to both domain knowledge and IS&R knowledge, the second to IS&R knowledge and the third to domain knowledge in Figure 2.11. The fourth infrastructure associates to contextual factors in Figure 2.9. The survey respondents have answered questions about both the context and self rated their search skills. The link between the personal information infrastructure, as described above and information literacy describes Marchionini as (Marchionini, 1999, "Information literacy")

> "Our personal information infrastructures are applied to information problems in an array of contexts and continue to evolve as a result of our struggles with and conquests of these problems. The development of our personal information infrastructure is roughly equivalent to our level of information literacy. Thus, information literacy is best considered to be a continuum of skills, concepts, attitudes, and experiences related to information access, understanding, evaluation, communication, application, creation, and value".

Marchioninis' approach to information literacy means that improved skills and knowledge leads to improved information literacy and therefore more efficient sub-processes during information searching. The thesis focuses on search skills and not information literacy as stated earlier, but search skills is used in a broader sense than Marchionini does in the quote above.

4.2.1 Modes of searching

Information searching can be of different kind. Bates divides information searching into four modes, and they are either active or passive and either directed or undirected. The different modes are: searching, browsing, monitoring and being aware. She estimates the active strategies searching and browsing to be a minor part of the users' total use of information seeking strategies, and that browsing is more widely used than searching (Bates, 2002). Within each of the strategies there are levels of search activities. The interaction can be divided into four levels: move, tactic, stratagem and strategy, where moves are the single actions during the interactions and strategy might stretch over long periods of time (Bates, 1990).

Traditional information retrieval is matching user queries with representations (records) of documents, and the model is called look-up retrieval (White & Roth, 2009). Web search engines, internal search engines and search functions in databases are based on the model. Other names of

this kind of searches are known-item search or analytic search (Marchionini, 1995). A well-defined query based on a well-defined information need is required for a successful look-up search. Exploratory search focuses on the kind of searches not emanating from a well-defined information need or a search for a known item, so called look up searches (Figure 4.2). Exploratory search is used for searching in the context of learning and investigating (Marchionini, 2006a).

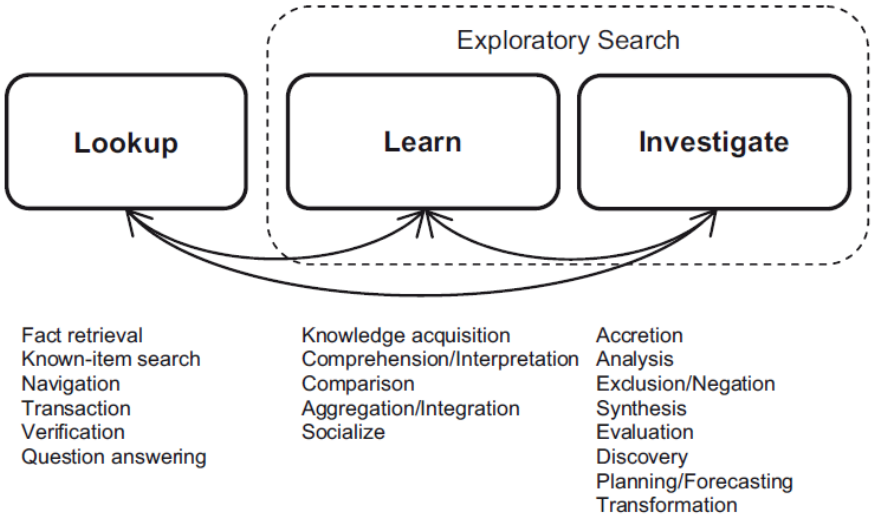


Figure 4.2. Exploratory search activities (White & Roth, 2009, p. 14).

As the search process proceeds in exploratory search the perceived problem becomes clearer and the level of uncertainty drops, as explained by Kuhlthau (1991). In the beginning of the search process the major tactic used is exploratory browsing, and later focused searching replaces browsing as the main tactic when the search process continues (White & Roth, 2009). A fourth type of search activity besides the three in Figure 4.2 is casual information searching as a part of a casual-leisure information behaviour which is motivated by hedonistic needs rather than information needs emanating from a work task (Elsweiler et al., 2011). The different search activities are used in the analysis of the survey findings (see Chapter 9).

4.2.2 Behaviour of web users

People use websites in some general manners. Users prioritize easy access, not quality, according to the principle of least effort. "Every individual when considering a course of action, will choose the action that requires the least amount of effort." stated Zipf (1949). In other words the users *satisfice*, they don't make optimal choices (Prabha et al., 2007). Another aspect of the principle of least effort is that web users muddle through, they do not figure things out (Ding & Lin, 2010). "Don't make me think" is the first of Krug's laws of usability (Krug, 2006).

Web users rely more on web search engines than on individual web sites. During a usability test the users went to a search engine in 88% of the time (Nielsen & Loranger, 2006). This stresses

the importance of being findable on the web, and the ranking in web search engines is crucial. Web users scan pages, they do not read them. First users scan the upper part of the content area in a horizontal way. Then they move down a bit and scan a second horizontal part. Finally users scan the content on the left side. These scans typical forms an F, a general, rough F shape (Nielsen, 2006).

The usage of the resources may be in many forms, from searching and browsing to reading and scanning. The term navigation is used in the thesisfor all forms of movement to and within the resources because it is neutral when it comes to intentions of the users. The users' behaviour can only be categorised in basic categories though the log files only contains behavioural data.

In a study of user behaviour on a museum web site Skov found four types of searching behaviour: Exploratory behaviour, highly visual experience, meaning making, and known item/element searching. The behaviours were mainly based on retrospective think-aloud sessions (Skov, 2009; Skov & Ingwersen, 2008). The different behaviours show that users or at least single search sessions are driven or motivated in different ways, for example finding the unexpected or by scanning the visual elements.

4.2.3 Navigational strategies on the web

There are three main information navigation strategies on the web (Levene, 2010; Nachmias & Gilad, 2002). The strategies are presented below, but number two has been reformulated to cover a broader, more up to date information search behaviour. Directories are not that important longer as they once were, instead social media sites and social bookmark services are used. The information navigation strategies:

Direct navigation. This is the simplest way of navigation is typing in the URL in the browser. This strategy includes other ways of navigation to known destinations like following a bookmark saved in the browser or a previously visited URL in the browsers search history.

Navigation from a site. Following links on website and blogs is the normal usage of the hypertext-nature of the web.

Navigation using a Search Engine. Search engines are used both for navigation the web and searching for information. "Traditional" search questions are not the only type of searches done in a search engine. A lot of the searches are to websites known to the user and the search engines are used as a fast tool for navigation.

These three strategies will be used in the broader sense of web navigation, not just information searching, as they cover the three ways of navigation in a hypertext. The use of the three strategies varies over time and depends on the development of the web. During the early years of the web, before search engines like Alta Vista and Google, where navigation through link directories like Yahoo, the most common strategy. In the 1990s, was still of information available on the web small, so the central parts of the web could be catalogued.

To reach the information objects the user must go through the web to the information system that contains the objects using one of the three strategies described above. The strategies are divided into external and internal strategies. The external strategies cover navigation on the web for information systems and files (Table 4.2), while the internal strategies (Table 4.3) are navigation within an information system for an object, such as within a site. The strategies divided into steps:

Table 4.2. Steps in external information navigation strategies (the letters correspond to the arrows in Figure 4.3).

Strategy	Step 1	Step 2	Step 3
Direct navigation (1)	Known object URL (a)	Object	
	Known resource URL (b)	Resource*	
Navigation from a site (2)	Follow link to object (c)	Object	
	Follow link to resource (d)	Resource*	
Navigation using a Search Engine (3)	Querying a search engine (e)	Follow link in SERP to an object (f)	Object
	Querying a search engine (e)	Follow link in SERP to a resource (g)	Resource*

* Leads to internal information navigation strategies (Table 4.3 below).

Within a resource there might be different possibilities of navigation, to follow the hierarchy of links and using the internal search engine. A resource may facilitate one or both of the navigational ways. An example of a resource which often only has search possibilities is OPACs, online library catalogues. Small websites on the other hand normally facilities navigation through links as the only mean of navigation.

Table 4.3. Steps in internal information navigation strategies (the letters correspond to the arrows in Figure 4.3).

Strategy within resource	Step 1	Step 2	Step 3
Navigation in link structure (4)	Follow link to category page [from home page] (h)	Follow link to object [from category page] (i)	Object
	Follow link on object page to other object page (l)	Object	
Navigation using internal search engine (5)	Querying internal search engine (j)	Follow link in SERP to an object (k)	Object

The steps in Table 4.2 and Table 4.3 above become more concrete when they are inserted in the model where the objects are embedded into the information and the web (Figure 4.3, using the model Figure 2.7b as foundation). Then it becomes clearer what the different navigation strategies are, how the user moves between the different layers to achieve the objects. Each arrow requires that the strategy is supported by the structure of the system to be effective in practice. Potential obstacles in the structure can be circumvented by the relevant search strategies along with situation-specific tactics. Search skills are knowledge of the use of different tactics and strategies in real situations, to further the search and to overcome obstacles.

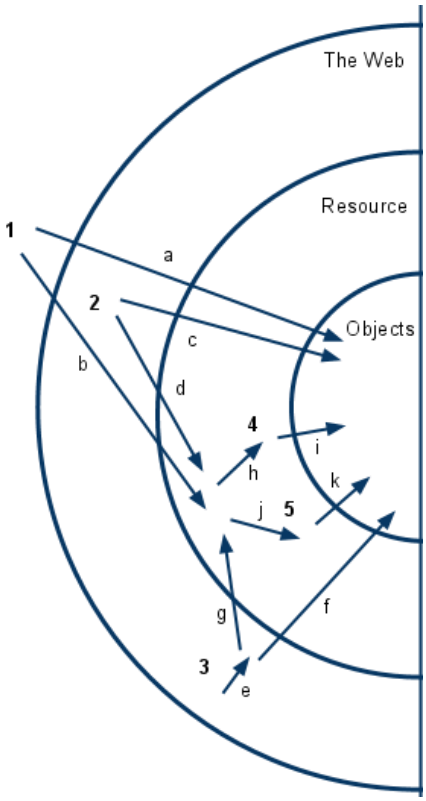


Figure 4.3. Illustration of the paths of the different external (1-3) and internal (4-5) navigation strategies in the object-model (Figure 2.7b).

The illustration above shows the five ways to the objects (arrows a, c, f, i and k). And the three ways to the resource are shown with arrow b, d and g. Arrows h and j shows the two fundamental different ways of internal navigation, browsing and internal search. The letters corresponds to a step in the different strategies (Table 4.2 and Table 4.3). The difference between strategy 1 on one hand and strategy 2 and 3 on the other is that direct navigation is able to take its starting point outside the web. Direct navigation can also be executed from any page on the web or from outside the web, in any moment the user is able to "teleport" to a known URL, through writing a known URL in the browser, choosing a bookmark or using the visited URL history in the browser. The different variants of the strategies are illustrated below in Figure 4.4, Figure 4.5 and Figure 4.6. To be extra clear they are pictured in both the object model (Figure 2.7b) and in relation to the resource model (Figure 2.7a).

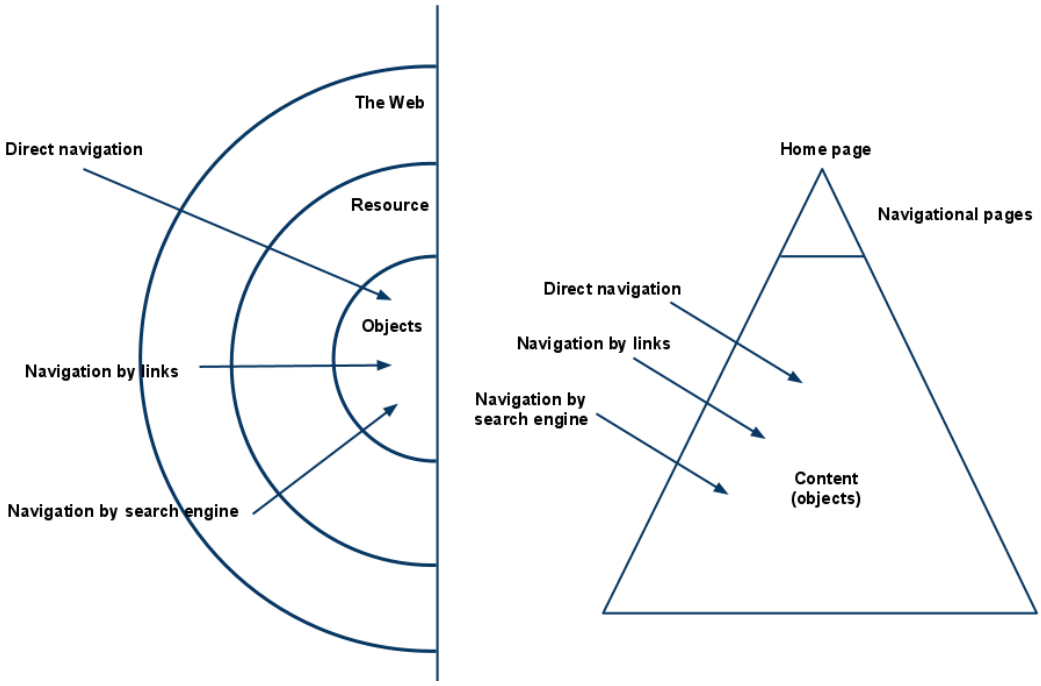


Figure 4.4. Navigation strategies with the end point at specific objects in the object model and in the resource model. Note that all the strategies bypass the resource level in the object-model.

The three navigation strategies that end at an object by passes, "jumps" over, the resource level, as illustrated in Figure 4.4. The content on navigational pages is bypassed and specific objects are reached directly.

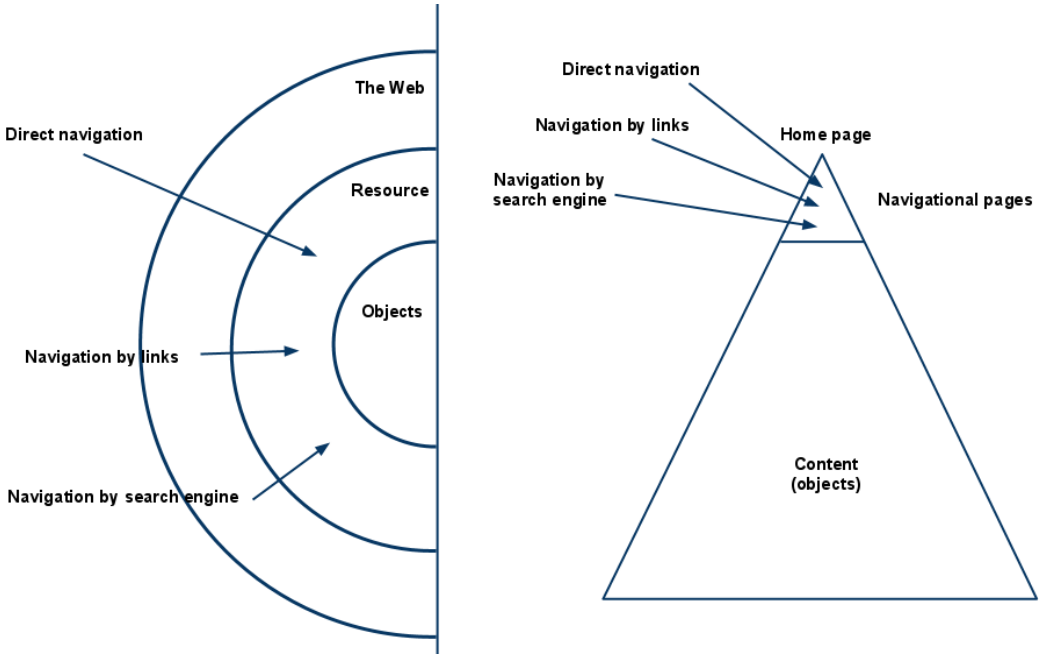


Figure 4.5. Navigation strategies with the end point at resource level in the object model and in the resource model.

In Figure 4.5 the three navigations strategies ends at the resource level and the user arrives at one of the navigation pages in top of the resource triangle. This means that the user is forced to use internal navigation ways, either the link structure or the internal search engine, to reach the objects (Figure 4.6).

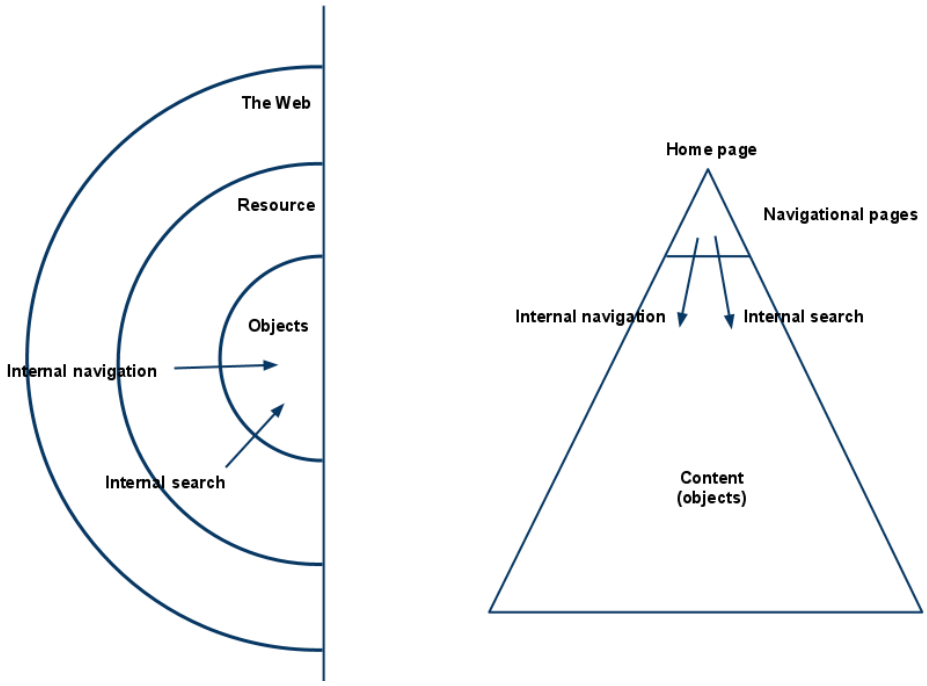


Figure 4.6. Navigation strategies with the starting point at the resource level and end point at specific objects in the object model and in the resource model.

Within a resource there are two ways on navigation, to follow links (browse) or to use the internal search engine (search). Normally one or the other is studied; either the focus is on the link structure and how users navigate through the structure, or the focus is on the utilization of the search engine and how users formulate or reformulate queries (Spink & Jansen, 2004). Single studies have looked at both ways of navigation and how users change way of navigation within sessions (Mat-Hassan & Levene, 2005).

User interaction with general web search engines like Google and Bing has been studied extensively, but normally just on some available, older, log files from Altavista and Excite (e.g. Spink & Jansen, 2004). In this research query formulation and reformulation has been in focus, together with factors like session length and media types. The studied query formulation has not been related to the information seeking process, the outcome of the search beyond the session in the search engine.

The Ciber research group has since 2000 studied the usage of different web resources with their deep log approach. In the *Deep Log Analysis* data from transaction logs is combined with data about the users, e.g. user profiles or the academic department the user is affiliated with (Nicholas, 2009; Nicholas et al., 2006a; Nicholas et al., 2006b).

4.2.4 Paths within the resources

As the resource model shows in the figures above, there are two basic arrival points in the resource, at some navigational page (N) or at a cultural heritage object (O).

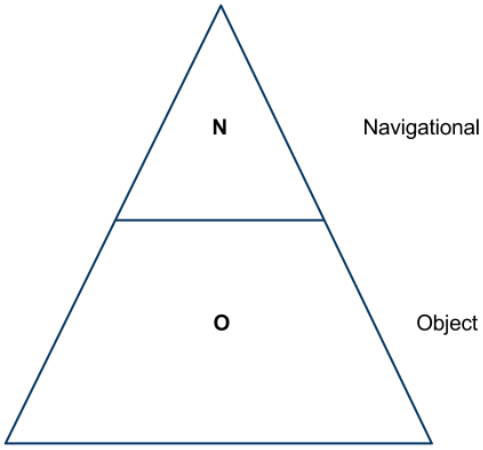


Figure 4.7. A simplified version of the resource model (Figure 2.7a).

In N there are then several possibilities in the way the session can be:

a. The session is just on page view long and the user never leaves the arrival point (in N), no cultural heritage object viewed.
b. The session continues from navigational pages (N) to the cultural heritage objects (O), and possibly up again (to N) and so on.
c. The session stays in the navigational pages (N), and the user never reaches the cultural heritage objects (O).

In the same way the sessions starts with arrival in O can be of different kind:

a. The session is just on page view long and the user never leaves the arrival point (in O), just one cultural heritage object is viewed.
b. The session continues from cultural heritage objects (O) to navigational pages (N) and possibly down again (to O) and so on.
c. The session stays among the cultural heritage objects (O), and the user never reaches the navigational pages (N).

Different navigation paths within the resource are illustrated in Figure 4.8 below. In the first two cases, N1 and O1, there is no path within the resource because the sessions are just one page view long. In the other four cases more than one page is viewed. In 1b and 1c the user arrives at the resource level, and then moves on in both levels (N3) or just in the resource level (N5). In O3 and O5 the user arrives at the object level, and then moves on in both levels (O3) or just in the object level (O5).

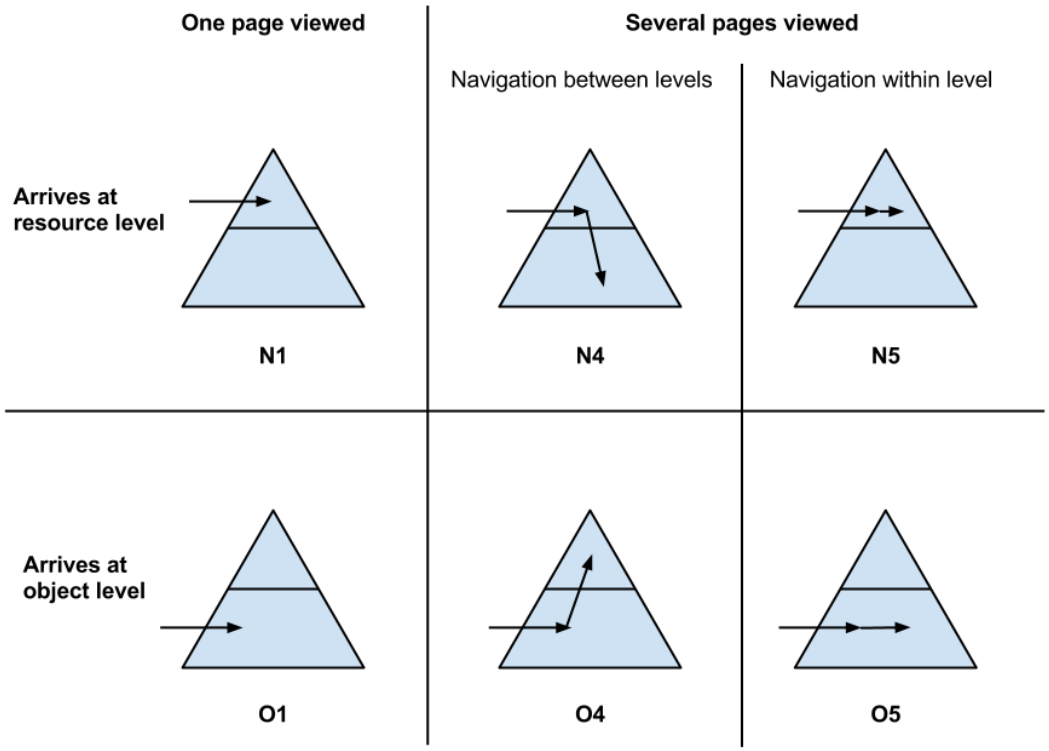


Figure 4.8. The six session paths in the simplified two level resource model. The second arrow in B4, N5, O4 and O5 represents one or several pages viewed at the level.

By looking at the paths in a more general way it is possible to study how, and if, the users' moves between the different levels within a resource. Together with the navigation strategies it is possible to study differences in the internal navigation based due to which navigation strategy the user used to get to the resource. This is also important to study the different paths because in N1 and N5 in Figure 4.8 no cultural heritage objects are accessed at all.

In the present study of the paths in Poma two levels are used in the site structure analysis, findability analysis and log analysis based on the structure of the resource. In the analysis of ADL and KID a third level is added between the navigation level and the object level, as shown in Figure 2.7. The triangular resource model (a) with the parts of a web resource: navigational objects, informational objects and cultural heritage objects. The circular object model (b) with the objects embedded in the resource, which is available on the web.. The path models, Figure 4.8, forms the framework for quantifying the use of strategies observed in the log analyses, Sections 5.4.8 and 8.2.3.

4.3 Chapter summary

To answer the question *What measurements or indicators of usage are central for measuring the navigation to the resources and the actual use?* aspects concerning the user side of the URI model (Figure 2.11) were discussed. Central pieces of relevant research on users, information needs, search skills, and information source horizons were discussed as background for the web survey. The focus of the chapter has been on web navigation in the context of information seeking. The three navigation strategies on the web are: (1) direct navigation through a bookmark or by typing the URL; (2) navigation by links; and (3) navigation by using a search engine. When the navigation strategies are combined with the object and resource models the important dimensions in the information system in relation the each strategy.

Based on the resource models number of levels there are a number of possible paths the user can take within the resource. The different session paths are used in the analysis quantitatively to study which kind of objects the users accesses, and in the extension how many of the visitors who actually accesses the digitalized cultural heritage.

A number of indicators from both the log analysis and the survey are used to study the usage and the users. In the log analysis the indicators are one of the following four kinds: Content usage; Referrals; User behaviour; and, Basic user characteristics. From the survey complementary indicators about the context and purpose with the present visit and the navigation strategy used will be gathered alongside characteristics of the users. These datasets will provide indications of the usage and the users, and together the indications will answer the second research question, *How do users find and use the cultural heritage resources?*

5 Methodology

The research design presented in this chapter is based on the conceptual framework in Chapter 2. The chapter answers the following questions: *How can an appropriate research designed be shaped to collect both usage data and findability data from cultural heritage resources?*

The first part of the chapter is about the research design, triangulation and mixed methods research. Then the chosen methods, site structure analysis, findability analysis, log analysis and web survey are addressed. The focus of the thesis is wide and methodology is complex. The chapter ends with a discussion about mixing methods and combining different kinds of data and indicators. The last section is a summary of all four theoretical chapters before moving on to the empirical part.

5.1 Research design

The traditional process is to formulate research questions and then choose best methods to answer the questions, in other words the research questions dictates the methods. In mixed methods research, which is discussed below, another view on the role of the research questions has been proposed: the research questions as the hub of the research process (Plano Clark & Badiee, 2010). The research question is placed in the centre and interplaying with the surrounding aspects: purposes; theories and beliefs; methods; and validity. The whole system of interaction is placed within environmental contexts, e.g. funding goals and research programs, which can also influence the research. The research question is still important, but the surrounding aspects are integrated in the research process and thereby with the research questions (Plano Clark & Badiee, 2010, p. 280).

Several perspectives on cultural heritage resources and their users are integrated in the thesis. The study is limited by some external demands, e.g. the use of several methods (Section 1.7). The research design is also determined by the resources and time available. The main determinator is the externally demanded focus on everyday life usage. The research questions have been developed parallel to the development of the conceptual framework and the explorations of the log files, not in advance. Besides the limitations mentioned in Chapter 1 three starting points were identified as central based on the objectives of the thesis (Section 1.2):

1. Log files from web cultural heritage resources are important data sources as they contain all types of access, both usage in work contexts and in everyday life (as pointed out by the ELIS framework).

2. How easy a resource is to find on the web determines how it is used. Hard-to-find resources drowns in the large amount of data on the web and most of the web resources competes with similar resources.
3. Information searching is an activity in which both the process and the outcome depends on continuous interactions. The user interacts with the information environment in an ecological manner.

The three starting points, discussed in the previous chapters, have influenced the research design. The conceptual framework does not limit which methods that is possible to use or the type of data gathered, despite it is founded in IIR. For example, in the present research design the survey and the findability analysis were chosen to supplement the log analysis. The methodological set up is explained in the next section. Several methods were rejected. Surveys with qualitative questions instead of quantitative were rejected due to the imagined difficulties of relating qualitative answers to the data derived from the log files, and because of the amount of work the analysis would demand. Interviews would have been an alternative if the research was focused on a particular group of users. For example, the participants in a study circle relevant for one of the cultural heritage resources could have been a good group to interview. The findability analysis could have been developed in another, more automated manner, but it would have required software programming skills.

5.1.1 The framework and the research design of the project

The research design in form of the different methods used and data analysed is illustrated in Figure 5.1. In the figure the arrows on each level are numbered and the numbers are used to describe different sub studies within the study.

The degree of findability (RQ1) will be studied by a findability analysis (arrow D in Figure 5.1) at the structural level. On the content level, some form of content analysis will be done (B in Figure 5.1). Arrow C in Figure 5.1, the feedback from the system, might be a part of the log analysis when it comes to internal search; otherwise the feedback is embedded in the hypertext navigation where each click on a link transfers the user to a new page with new content and structure.

RQ2: How do different groups of users find and use the cultural heritage resources?, will be answered with data about the user aspects (F, G and H in Figure 5.1). The main empirical data in the log files from the cultural heritage resources where the actions of the users will be explored (G in Figure 5.1). In addition to the log file analysis web surveys' will be completed to gather data about the users' information needs (H in Figure 5.1) and their level of information search skills (F in Figure 5.1). The survey will also gather background data about the users' age, gender, level of education, among others. Some aspects of the users' intentions will be part of the log analysis as search terms referral URL's, which will be analysed with Broder's taxonomy of web search (Broder, 2002).

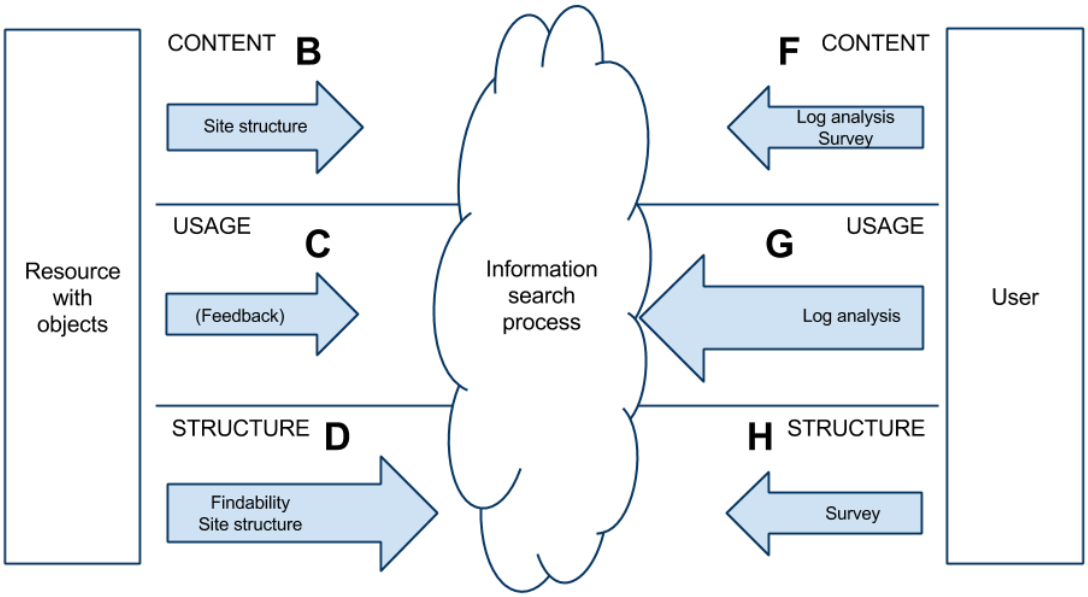


Figure 5.1. The URI model (Figure 2.11) is the conceptual framework the research design is built upon. Here with weighted size of the arrows based on their importance in the research design. The letters are the same as in Figure 2.11.

To answer RQ2 about usage transaction log analysis is chosen as primarily method (G in Figure 5.1). Log analysis is chosen for two interlinked reasons. First, it captures real user interactions based on real information needs. Second, the method captures all usage; no group of users is studied specifically. The log files contain the usage of the resources. The everyday life usage is hard to capture in other ways without leaving the focus on the resources. It is possible to study a specific group of users, e.g. how hobbyists interested in artwork, uses a cultural heritage resource, but then some of the focus will be on the specific group of users and not the resource and the other users. The study of information behaviour in everyday life is a challenging task, and each method used is accompanied with disadvantages. In this case the advantages with studying everyday life usage as a part of all usage with log analysis was seen as a reasonable trade-off, in all other possible methods just a minor part of the everyday life usage would be studied. In addition to log analysis web surveys are used to collect more data about the users and the why they are visiting the resources (primarily F and H in Figure 5.1), due to the reasons explained by Jansen:

> "Surveys gather data on respondents' recollections or opinions; therefore, surveys provide an excellent companion method for Web analytics that typically focus exclusively on actual behaviors of participants" (Jansen, 2009b, p. 51; based on Rainie & Jansen, 2009)

Combining log analysis with web surveys makes it possible to gather a different kind of data than the actions of the users captured in the logs. There is no alternative to the survey for collecting quantitative data on a large group of users concerning demographic data, search context and intentions and aspects missing in the logs.

On the system side the findability of the resources and their objects are measured/evaluated through a findability analysis (D in Figure 5.1). The basic structure of the resources is studied through a site structure analysis based on the resource model, Figure 2.7 (B and C in Figure 5.1). The site structure analysis is used as a support method to both the log analysis and the findability analysis.

5.1.2 Measuring the information search process

As discussed by Boyce et al. it is the process of searching, or interactive information retrieval, which is studied through the log files, not the outcome of the process.

> "The process measures do not deal with outcome. They do not deal with any of the various aspects of the user's satisfaction with the results of a search. Process is largely concerned with mechanical actions." (Boyce et al., 1994, p. 168)

The process of IIR is the sequence of events that together constitutes a session (Boyce et al., 1994). Boyce et al divide process measures into two broad categories: "(1) measures concerning *direct monetary cost* and *use of resources* and (2) measures concerning *time* and *numbers of commands* issued*"* (Boyce et al., 1994). The distinction is made in the context of database search, but if they are rephrased as measures concerning: (1) *attentional cost* and *use of resource* and (2) *time* and *numbers of actions* taken; they are applicable on web resources. For free resources on the web there is no monetary cost, the cost is the attentional focus and time spent by the visitor. The cost may be measured as the number of clicks in a session:

> "Since the values found by a user in the pages she visits while surfing constitutes a random process, even a frequent user visiting the same site will go through a different number of clicks in every session. Thus, the only meaningful quantitatives to speak of are the average number of clicks per session, as opposed to the exact number that a person will go through at a particular time in a given day." (Huberman, 2001, p. 45)

But, at the same time when studying web navigation which tend to have statistically skewed distributions "the average conveys little information on surfing patterns" (Huberman, 2001, p. 47). This particular problem will be handled by separating the frequent one-page-viewed sessions as one category of session paths.

Hung et al. (2008) has illustrated the observable moves or actions in their model of contextualized information searching (Figure 5.2). Information searching is divided into four levels, ranging from the overall *grand strategy* on the upper level, to *strategy* (level 2) and *tactics* (level 3), and finally *operations* on level 4. The hierarchy is similar to Bates (1990): *strategy*, *stratagem*, *tactic* and *move*. The users operates on a cognitive level on the first three levels in the model. It is only on the lowest of the four levels that the actions or moves of the user are observable, and those actions constitutes the session together with feedback from the information system or change in the information space due to the user's interaction with the system.

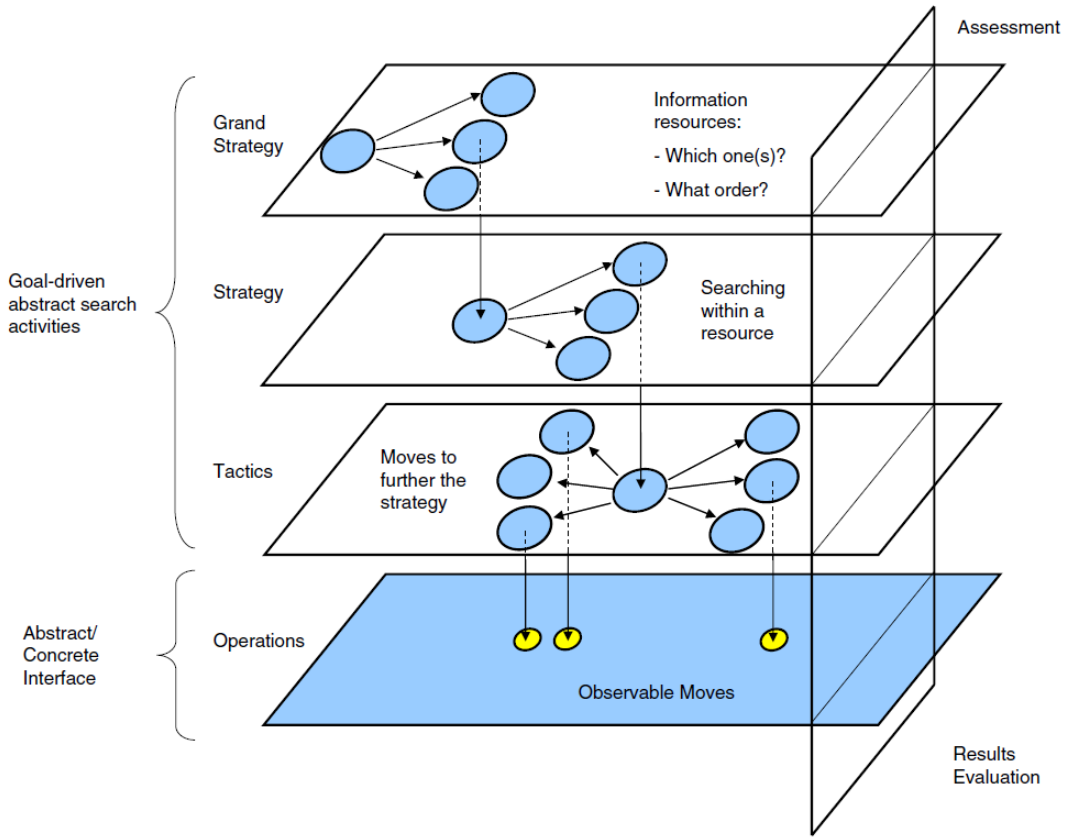


Figure 5.2. A multi-level model of contextualized information searching (Hung et al., 2008). The three upper levels are abstract in nature and the lowest fourth level, which is shaded blue, is where the concrete, observable actions of the users takes place. The observable moves (the yellow dots on the fourth level) forms the measurable search process.

The model above highlights the important fact that just the observable moves on the operational level can be studied through observation, regardless if the method used is eye-tracking, video recording or log analysis. Complementary methods, such as survey or think aloud can be used to capture some of the activities on the abstract goal-driven levels, i.e. the parts of the search process that are the hardest to measure. The consequence of the limitations of observation is the use of triangulation, which is discussed below.

5.1.3 Triangulation and mixed methods research

Triangulation is the practice of using different positions, points of view, to study an object or problem. The main goal is to get a better understanding of the studied problem. The notion of triangulation is based in trigonometry and geometrics of triangles. If two corners of a triangle are known the position of the third corner can be calculated. In social sciences it is used in a metaphorical sense, not in an exact mathematical way (Denscombe, 2010). According to Denzin there are several types of triangulation in social science research. He discusses four kinds of triangulation: data, investigators, theories and methodologies (Denzin, 1970).

Methodological triangulation can be both between-methods and within-methods. In the case of between-methods triangulation different methods are used, for example a quantitative survey and a qualitative interview may be used in combination. The markedly different methods give the researcher two views that are far apart. The within-methods triangulation is used for another purpose, to evaluate and analyse quantitative data and the develop research instruments, e.g. using a standard personality test when testing out a new one. If two similar methods generate the same results it is more likely that the findings are accurate and authentic, and not generated as a by-product of the method used (Denscombe, 2010).

Data triangulation can be used to check the validity of findings. The different data may be collected from different sources, at different times or in different spaces. Another form of triangulation is investigator triangulation where the findings of different researchers are compared for consistency. The last form of triangulation according to Denzin is theory triangulation. Here more than one theoretical viewpoint is used in relation to the data, both in terms of collecting the data and analysing it (Denscombe, 2010). Another distinction between different kinds of triangulation is done by Turner and Turner (2009). They talk about hard and soft triangulation. Hard triangulation is when triangulation is used to challenge and test findings, where as soft triangulation is confirmatory and complementary.

There are two main benefits of using triangulation, e.g. two different approaches to triangulation, according to Denscombe. The first is improved accuracy where the alternative methods are used to confirm the accuracy and authenticity of the findings. The second approach is a fuller picture, completeness, of the findings by combining different facets or build up the research by employing different methods at different stages (Denscombe, 2010). With several methods in the investigation the confidence in the conclusions might be seen as higher (Bryman, 1988).

On the other hand Denscombe point out three drawbacks which have to be weighed against the benefits when using triangulation. First, the use of several methods will limit the use of the single methods deployed in depth and scope, and it demands multiple skills. The second drawback is that the data analysis becomes more complex when using triangulation. Several kinds of analyses have to be completed and the findings has to be compared, contrasted and integrated in more demanding ways. The third drawback or risk is that the different findings gained as a result of the triangulation do not support each other. In the long run, Denscombe means this should result in further research, but in the short run this might be a problem as research tasks have to be completed (Denscombe, 2010).

Triangulation within social science is mainly discussed in the tradition of mixed methods research. Mixed methods research (Mixed research) is based on "the belief that treating qualitative and quantitative approaches to research as incompatible opposites is neither helpful nor realistic when it comes to research activity" (Denscombe, 2010, p. 139). Based on published reviews on mixed methods research Denscome concludes that mixed methods are used for different purposes (2010, pp. 139-141):

1. improved accuracy

2. a more complete picture
3. compensating strength and weaknesses
4. developing the analysis
5. an aid to sampling

Denzin, the early advocate of triangulation, is "against" the present use of triangulation and mixed methods research. Triangulation in (Denzin, 1970) was only intended to be used on qualitative methods (Denzin, 2012). The relation between triangulation and mixed methods research remains unclear, at least within the mixed methods research community. Bergman highlights three possible relationships: (1) triangulation is a subset of mixed methods research; (2) triangulation and mixed methods research are synonyms; and (3) mixed methods research is a subset of triangulation (M. M. Bergman, 2011). It is also disputed if triangulation with only qualitative or quantitative methods is mixed methods research (M. M. Bergman, 2011). The use of triangulation and a mixed methods approach is pragmatic. As presented above neither triangulation nor mixed methods research is unambiguous. The main advantage of the approaches is that they have a non-black-and-white view on qualitative respectively quantitative research. Both research traditions have valuable strengths, and the present research have features from both quantitative and qualitative research.

In the research design different kinds of triangulation is deployed. Method triangulation (between-methods) which combines the results from all the four methods, log analysis, web survey, site structure analysis and findability analysis, is used to get a fuller picture. But the methods are also combined to compensate for their strength and weaknesses, e.g. the web survey complements the log analysis in terms of user background. Data triangulation is practiced by studying three different cultural heritage resources.

For RQ1 a method for findability analysis is developed as detailed in Chapter 3. For RQ2 the main method is log file analysis, to gather navigational data. The goal is to answer:" What do the users do in the resources?" The behavioural data from the log files is complemented with web surveys, to get a picture of the users (who uses the resources and why they are used).

The site structure analysis precedes both the analysis of the logs and the findability analysis as they both uses the results of the site structure analysis, a sequential mixed design according to Teddlie and Tashakkori. The survey is deployed parallel to the three first methods, a parallel mixed design (Teddlie & Tashakkori, 2009). The methodological triangulation is illustrated in Figure 5.3.

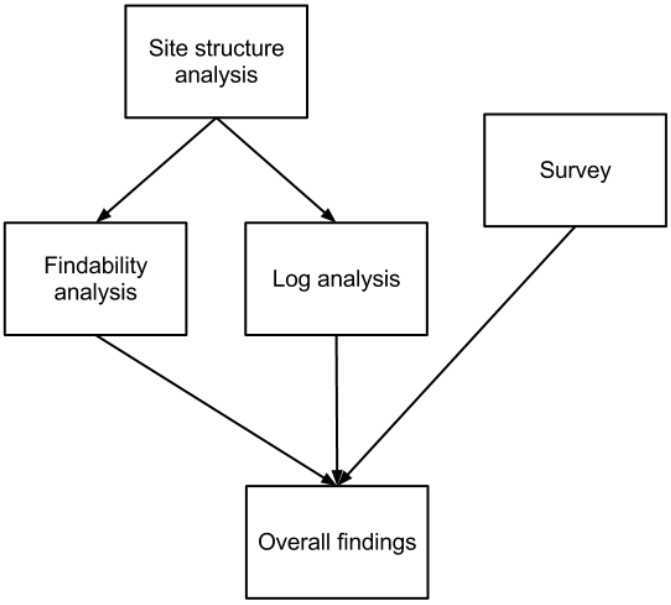


Figure 5.3. The methods form a methodological triangulation, here presented in relation to each other. The survey is parallel to the other methods, and the site structure precedes the findability analysis and the log analysis in the mixed methods setup.

5.1.4 Usage and user data

It is possible to extract or calculate different usage indicators from log files. The possibilities are determined by the format of the log files, and the parameters that are stored in each transaction (see Section 5.4.1). The indicators are often calculated per time unit, e.g. month, per visitor, or listed after popularity. Examples of usage-indicators are: Number of sessions per month, number of page views per session, or most popular page. The indicators can be divided into different categories. One example is: *Site usage*, *Referrers*, *Site content analysis*, and *Quality assurance* (Jansen, 2009b, p. 30). Another categorization is the one based on the *Deep Log Analysis approach* from the CIBER research group: *Activity metrics [indicators]*, *Information seeking characteristics* (including both *Type of content viewed* and *Searching style*), and *User characteristics* (Nicholas, 2009, p. 123). As can be seen, the first categorization has a website focus, in the Web analytic-tradition. The second one is focused on the user. In the Deep log analysis approach the log files are complemented by other data about the users so the analysis can be deeper. The Deep log analysis research is done in the traditions of *Transaction log analysis* and *Human Information Behaviour*. Some indicators can advantageously be based on log files:

- Content usage (number of sessions, most popular pages, etc.)
- Referrals (navigation strategies, search terms used in referring search engine, etc.)
- User behavior (paths in resource, bounce rate, etc.)
- Basic user characteristics (location, web browser used, etc.)

Other indicators are not or only partly covered in log files and that is the reason why the log analysis always is complemented with other data types in the Deep log approach for a richer data set:

- Information need or intention behind search
- Users' level of search skills and domain knowledge
- Contexts of the sessions
- Demographic data about users

For research in information seeking and retrieval some indicators are more important than others. There are several standards for web log files, but what is stored and how is always possible to change by the site owner. Basically all logs contains the URL of the requested page, the time of the request, the URL of the referring page (where the user clicks on the link to the requested page), and the IP number of the computer doing the request. Together these four factors can be combined into different indicators. The first one is straight forward, the *page view*. The page view indicators are based on the URL request together with time data and typical indicators are: most popular pages, top entry pages, and top exit pages.

Session is an indicator based on a series of page views by a single user. The session is set by the time of the requests and IP number of the request (sometimes together with a session-id or a session cookie). Page request close in time from the same computer is seen as a session. Normally the sessions are regarded as being ended by an time inactivity, of e.g. 30 or 60 minutes with no new page requests. A session can contain from one page view to hundreds or thousands of page views, but when it comes to really long sessions (both in time and number of page views) they are normally not due to human users. Often the long sessions are crawls by search engines as the search engine crawlers' visits or revisits many web sites frequently, and they often forms more than 50% of the traffic to a site as reported in Eurpeana (Clark, 2011).

The length of a session can be measured in two ways, in time or in number of page views. Time is a problematic measure because the only way is to compare the time-stamp of the second page request with the time-stamp of the first page request, and so on. Measured in time a single page view session has no time. Measured in page views the session length is equal to the number of page views within the session time out time. Within each session there is a *path*, the series of pages viewed. The paths can have different depth, or site penetration.

Another important indicator is the *referrer*. The referrer is the URL from the external web page in the first request in a session. It is the link in another site the user followed to get to the present site. If the referrer is a web search engine the search terms used in the search engine is normally included in the referrer, which is an indication of the users' initial need or intention. Basically there are four types of referrers: web search engines, web sites, direct and unknown. The first three corresponds with the navigation strategies on the web and in the fourth type, unknown; there is no information about referring URL in any of the page requests in the session.

User-indicators can only be derived from the IP number, together with session-id or session cookie if the technologies are in use. The geographic region or city of the computer in use can often be determined from the IP number, and sometimes the company or institution it belongs to. If the user was logged in to the web site there may be more information about the user in the user profile to make use of, as done in the Deep log analysis approach.

To measure returning visits some other data than just the IP number is needed because an IP number can refer to a shared computer or to a whole organization which uses a single IP number for the traffic outside their firewall.

A number of usage-indicators are used in the study. Some of the indicators are compounds of several indicators. Bounce rate is the percentage of sessions with only a page view relative to all sessions. In the thesis the usage analysis is based on the indicators described below. The indicators will illustrated in the resource model (Figure 2.7a), as for example in Figure 5.10.

- *Levels visited in session* (based on the site structure analysis in the triangular resource model) is used to study how many of the users actually looks at CH objects.
- *Session length/arrival level* to study if there is a difference in the length of the sessions based on where in the resource the sessions starts.
- *Session length/navigation strategy* to study if there is a difference in the length of the sessions based on how the users arrive to the resource.
- *Arrival level/navigation strategy* to study if there is a difference in arrival level between the different navigation strategies used.
- *Bounce rate/arrival level* and *bounce rate/navigation strategy* are used to study the bounce rate, the number of one page sessions divided by the total number of sessions, per arrival level and per navigation strategy.

The survey questions were designed to complement the data gathered through the log analysis.

- *Search task context* which captures in what context the CH resource is visited, e.g. for work or leisure.
- *Type of search task/intention with visit* is used to study type of search task in relation to the intention with the visit.
- *Navigation strategy used in visit*, which of the three navigation strategies was used.
- *Level of web search skills* is used to study the IS&R knowledge of the users (self rated).
- *Age* of the respondents.
- *Gender* of the respondents
- *Location* is the users present country of origin.
- *Education level (years of formal education)* of the respondents.
- *Present position* for capturing the current occupation or if the respondent is a student, retired, or unemployed.

The indicators cover the aspects discussed in Chapter 4.

5.2 Site structure analysis

How a resource is built up is often unique, but most resources have common traits. Every resource on the web have at least two levels, with one acting as entry or top level linking down to objects on other levels. These logical levels can be used to describe both the resource and the user's way through it. Normally the objects are more general in the top of a resource and the objects becomes more specific further down in the resource.

The site structure analysis is an analysis based on function, a combination of content and structure; two of the three webometric layers (see Section 2.6). Besides the use in describing the resources the result of the site structure analysis is used in both the log analysis and the findability analysis for differentiating the analyses. The structure analysis is discussed on a general level, and the results of the analysis of the three cultural heritage resource are presented in Chapter 6.

5.2.1 Determining levels of the CH resource

All information on the web is published in some sort of information system, as discussed in Chapter 2. How findable the information is depends on many things, but the user always has to find her way on the web to the resource containing the information objects. The model in Figure 2.7a. There are three basic levels in the resource, at some navigational page (N), at some informational page (I), or at a cultural heritage object (O) (as discussed in Section 2.3). This division is important when looking on search strategies and navigational ways on the web (see Chapter 4).

The site structure analysis is an interpretative process where the decisions taken influence the whole study. The KID structure in Figure 6.2 might serve as a example. Hypothetically, if the information from Weilbachs Kunstnerleksikon (the extended artist information available for some artists) is placed at the Object level instead of the Informational level in the site structure analysis, the results from the following analyses will be different. The distribution of paths would change, as well as the average session length and bounce rate (per level). Also the results of the findability analysis would be changed if the objects have lower or higher findability scores than the original objects in that level. However, this hypothetical example also requires that the objects about the artists are the purpose of the resource, not the digitized objects as in the resources studied in the present thesis.

The site structure analysis and its impact are deeply embedded in both the conceptual framework (Chapter 2) and the research questions. In the cultural heritage resources studied in the present study the levels are relative easy to distinguish: general information, information about the cultural heritage, and the cultural heritage itself. But it might be harder to classify all the pages of

a general web site into a couple of levels. For an e-commerce site it might be possible to see the check-out as the last step in a session path because it is the site owners goal that users (customers) go through with their purchase.

5.3 Findability analysis

The goal of the findability framework is an overall findability analysis, an operationalization of the theoretical findability framework discussed in Chapter 3. Besides the present implementation of the findability framework there are other ways to do the operationalization, see for example the discussions in Section 7.6 about different weightning and Section 7.7 about automatization.

5.3.1 Evaluating findability

As discussed in Chapter 3 the method chosen for measuring findability is structured observations with findability protocols, which generates primary data. A number of typical objects on each level in the three studied resources were chosen to represent, in a non-statistically manner, the objects on the level. The evaluated objects were awarded points on each aspect (see Section 3.3) that lead to different scores for total, external and internal findability for each object. The scores are on an ordinal scale. This method does not provide an absolute findability score, but an approximation that is useful for relative comparison or finding weak spots in a given resource.

The findability measures have different characteristics which reflect the impact the aspect has on findability. An increased amount of metadata (number of SAPs) in *Object attributes* increases the findability, and hence the Object attributes are additive. Any problems with the *Accessibilty* decreases the findability, and it is therefore subtracted. *Internal search* or *Internal search* are requirements for internal findability, there must be at least one of the aspects present for internal findability. *Reachability* is a requirement for findability on the web; if an object cannot be reached it cannot be found. *Web prestige* is additive, i.e. increased web prestige increases the findability. These different characteristics are reflected in each of the findability measures below and how they are weighted in relation to each other. The findability aspects that are requirements (internal navigation, internal search, and reachability) are given one point if the requirement is met, otherwise zero points. The other aspects (object attributes, accessibility, and web prestige) are weighted higher and are given up to three points each. They are not requirements, but all three are considered to have great influence on the findability when the basic requirements are met in the other aspects, and that is why they are worth more than one point each. The purpose is to construct a composite indicator findability indicator that takes the many possible aspects into account, see Table 5.8 as the result of this analysis. The points given in this first version of the findability evaluation framework are arbitrarily chosen, but they are set in a attempt to add weight to some aspects. The three weighted aspects are rated on a four point scale but could equally be

three or five point scales. The findability aspects could be weighted in other manners as discussed in Section 7.6.

5.3.2 The findability measurements

One *findability* indicator for each of the six findability aspects, except for *Object attributes* which is evaluated by two indicators, were identified in Section 3.3 (Table 3.2). The chosen indicators are aggregated from factors found in the literature in combination with the conceptual framework. In SEO the work on increasing findability is done on the micro level, a typical factor is the number of keywords in the title of the html page or in a heading (e.g. Walter, 2008). In the following section all the indicators are discussed, and the operationalization for the findability analysis is presented.

Concerning *object attributes* each studied object is evaluated according to the two findability measures in Table 5.1 and Table 5.2. The number of SAPs is counted and the object is given the corresponding number of points. If a text object is in full text the object gets an additional point .

Table 5.1. Findability measure and points for evaluation of objects attributes in the form of SAPs.

Number of SAPs	Points
None (0 SAPs)	0
Few (1-10 SAPs)	1
Many (11+ SAPs)	2

Table 5.2. Findability measure and points for evaluation of objects attributes in the form of full text.

Full text object	Points
No	0
Yes	1

For the object attributes the two measurements captures two central aspects of the object, the amount of metadata (in the form of SAPs) and if the object is full text or not. The types of the SAPs are not taken into consideration. In a more elaborated findability-framework it would be possible to distinguish between for example SAPs in headings, anchor text, and body-text, and the different types of SAPs could be weighted depending on the prominence of the type in accordance with the SAP indexing approach (Wormell, 1985). The length of the full text could also be measured, not just categorized as yes or no as in the present study. It could also be of interest to measure how frequent the SAPs occur in the object.

The *accessibility* evaluation is done by testing the compliance of the objects against WCAG 2.0 guidelines in an online WCAG-error-tester. Each studied object is evaluated according to the findability measure in Table 5.3. The result of the compliance to the WCAG 2.0 is categorized as one of four categories: *To many* (no compliance); *Many*; *Few*; and *None* (full compliance). The borders between the different categories are subjective and the results largely depend on how serious the errors are and the number of different errors. The URL of each of the studied objects were tested in AChecker (see below) which analysed according to the guidelines in WCAG 2.0

and reported the number of errors. The errors in AChecker are reported as known problems, likely problems, or potential problems. Every time an error occurs is reported as a problem, e.g. a missing text attribute can generate hundreds of problems on a single web page as it is frequently missing. The number of reported problems as well as their type was noted, and the seriousness in terms of findability loss was estimated based on the description of the problem according to the WCAG 2.0 guidelines. The number of errors, the type of errors and their assessed impact on findability were combined together into the accessibility-measure in Table 5.3.

Table 5.3. Findability measure and points for evaluation of accessibility.

Number of WCAG-errors	Points
To many (no compliance)	0
Many	1
Few	2
None (full compliance)	3

The testing site AChecker[21] is available on the web and is possible to access through an API (see Figure 5.4 for an example of a result page).

[21] http://achecker.ca

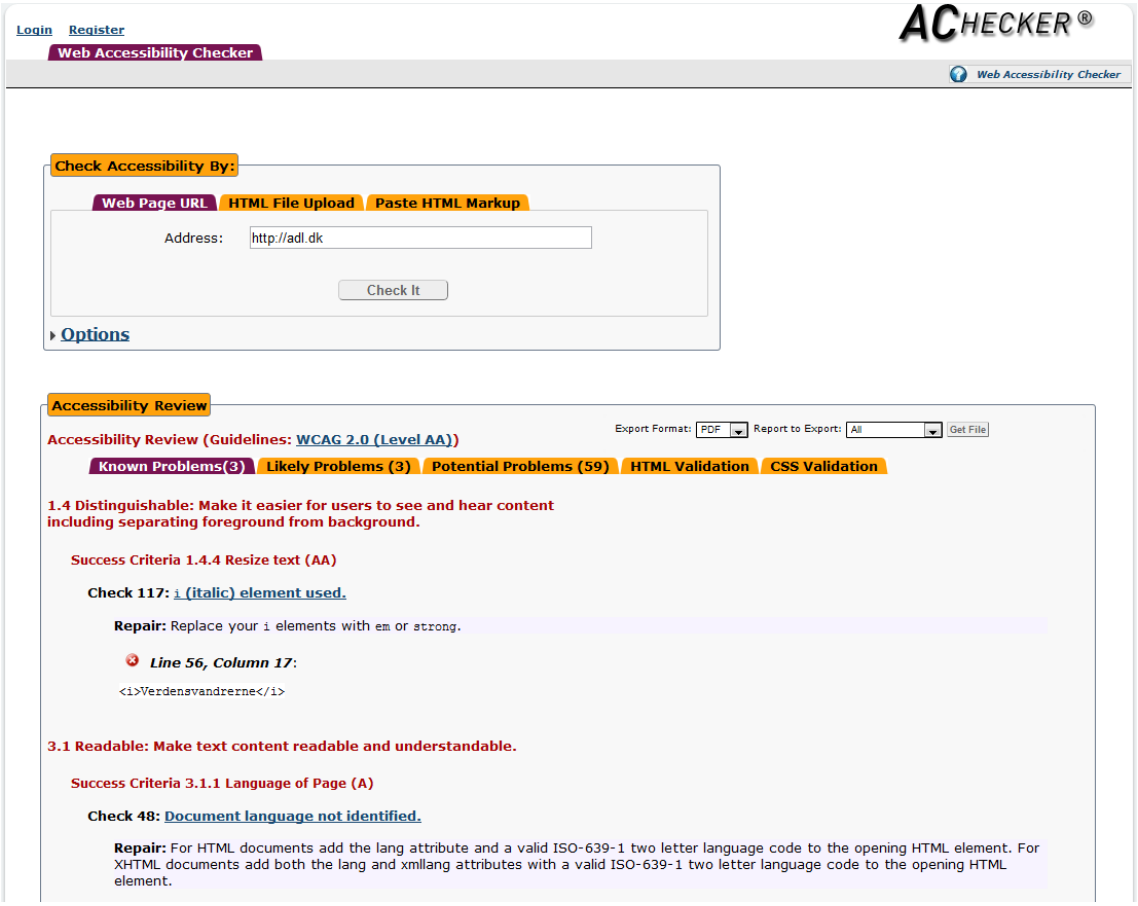


Figure 5.4. The Achecker (http://achecker.ca) test result for http://adl.dk. The number of problems are displayed in the tabs in the middle of the screen dump, e.g. "Known Problems (3)".

The result for each object was interpreted in terms of impact on findability. Only one tested object did not generate any errors or problems, the unformatted ASCII text in ADL. All other objects were reported having some problems, but none that had serious impact on the accessibility in the findability sense.

An alternative to the accessibility testing by testing against the WCAG 2.0 guidelines is to use a search engine spider simulator. A search engine spider simulator[22] is a way to test how search engines looks at sites and which information they can extract. The main difference, a part of their different focuses, is that the accessibility tester reports problems of page compared to an optimal page whereas the spider simulator just report what it can see, not what it misses. The issues of the page tested have impact on the results in both services, but it is harder to evaluate the magnitude of the problems in the spider simulator.

Internal navigation is about how and it is possible to find if an object through the links within a resource. In KID for example the number of actions (clicks) from top level to object is minimum

[22] For example http://www.webconfs.com/search-engine-spider-simulator.php

six actions. Five actions if the title had been presented on the first page of titles, in this case if the title had begun with A. Additional actions are needed, e.g. scrolling down or choosing a category letter, if the author was not presented on the top of the author list.

The evaluation of internal navigation does not take the visibility or prominence of specific objects into account nor the impact of internal visibility upon general internal findability. The CIBER research group has shown that if a web page is made more visible, the use of it increases (Huntington et al., 2004). But in the present study the overall findability is studied, and therefore the internal visibility is excluded as it often is the result of temporary promotion efforts of a specific object with a resource. The internal navigation is measured by studying if an object is possible to find through the internal link structure (Table 5.4).

Table 5.4. Findability measure and points for evaluation of internal navigation.

Possible to follow links to object	Points
No	0
Yes	1

Internal search is about if and how it is possible to find an object using the internal search engine. In KID for example it takes four actions using the internal search to find a specific text from the first page. It can also be about how high an object is ranked in the SERP. The relevance ranking depends on the query and is theoretically a query dependent measure.

Internal search is used as an indicator for the internal search aspect of findability. The complexity of studying the ranking of single objects in the internal search engine would be too great, and it would demand simulated or artificial queries. Simulated queries are a query dependent aspect and thereby not part of the present study. An evaluation with simulated questions would also take the object attributes into account when ranking is calculated, and those aspects are evaluated in other ways in the present findability-evaluation-framework. The internal search is measured by studying if an object is possible to find through the internal search engine (Table 5.5).

Table 5.5. Indicator and points for evaluation of internal search.

Possible to find object through internal search engine	Points
No	0
Yes	1

Reachability concerns whether it is possible to reach an object directly on the web - i.e. if the object has an URL which makes it possible to link to. The reachability is crucial for gaining web prestige (accumulate inlinks), and to facilitate indexing by the web search engines. The reachability is measured by studying if an object is possible to link to.

Table 5.6. Findability measure and points for evaluation of reachability.

Possible to link to object	Points
No	0
Yes	1

Web prestige is a measure which concers the authority or popularity of an object. The number of inlinks and the prestige of the inlinks decides the web prestige of the object measured. The only easy way to measure or evaluate web prestige is to use Toolbar PageRank as an approximate value of the web prestige of a web objects. PageRank is presented in Google Toolbar as a value between 0 and 10 and the value is a logarithmic simplification of the real value Google uses. A PageRank of six or above is considered high in a non-English context. If the resources were in English a PageRank value of at least seven would have been required to be seen as high (Westergren, 2009). Then, probably, five categories would have been needed to capture the whole scale of web prestige, and not four as in Table 5.7.

Table 5.7. Indicator and points for evaluation of web prestige.

PageRank-value	Points
No value (not ranked, might be indexed)	0
Low (PR 0-2)	1
Medium (PR 3-5)	2
High (PR 6-10)	3

There is no other public web link analysis value than PageRank, even if the competitors of Google might do similar calculations. For academic webometric research it would be of great use to have an open source web search engine (Bar-Ilan, 2004; Thelwall, 2012). PageRank is probably the most transparent part of Google search as it was developed at Stanford University before Google was launched (Brin & Page, 1998; Page et al., 1999). The PageRank values were collected by using the Google Toolbar in Internet Explorer. In the toolbar it is possible to show a PageRank-meter and by hovering the mouse over the meter at a given web page the PageRank value is displayed (see example in Figure 1.1).

Figure 5.5. An example of the PageRank meter in Google toolbar for Internet Explorer. The page about page in ADL has a PageRank of five out of the maximum ten.

5.3.3 Calculating total, external and internal findability

As discussed in Section 3.3.7 the total findability indicators are calculated differently. The external findability is calculated as the sum of the point given in the evaluation process described above to the four aspects that pertains the external findability: object attributes, accessibility, reachability, and web prestige. The score for external findability can range from zero to ten. In

the same manner internal findability is the sum of: object attributes, accessibility, internal navigation, and internal search. The score for internal findability can range from zero to eight. The findability measurements included in each type of findability are summarized in Table 5.8.

Table 5.8. Findability measurements based on aspect and level.

Aspect	Measurement	Score	Total findability	External findability	Internal findability
Object attributes	Level of SAP (a) Full text (b)	0-3	X	X	X
Accessibility	Compliance to WCAG-test	0-3	X	X	X
Internal navigation	Object linked to in internal link navigation	0-1	X		X
Internal search	Object indexed in internal search engine	0-1	X		X
Reachability	Stable and unique URL	0-1	X	X	
Web prestige	PageRank value of the object	0-3	X	X	
Maximum score			*12*	*10*	*8*

The maximum score for each type of findability is displayed in the bottom row in Table 5.8. The total findability score is not the sum of scores of external and internal findability as the first two aspects are included in both. For comparison the scores will be normalized and be expressed as a percentage of the maximum score, i.e. a total findability score of 9 will be presented as 75% of the maximal score (the score 9 divided by 12, which is the maximum score). This is inline with the nCG-measure for relevance (normalized Cumulative Gain) introduced by Järvelin and Kekäläinen (2002). This provides an opportunity to observe how far away from an ideal performance the analysed objects are. This makes it possible to compare the performance of several resources. The questions about the awarded points and the findability scores are also addressed in Chapter 7 after the findability analysis, in Sections 7.5 and 7.6.

5.4 Log analysis

There are two fields working with indicators from logs, *Transaction Log Analysis* and *Web Analytics*. Web Analytics is a new field that has evolved since the birth of the Web in the 1990s, mainly as a professional field. Transaction Log Analysis on the other hand has a long history in LIS research, the method has been used since the 1960s (Peters, 1993). The two concepts are overlapping to some extent. In Web Analytics web logs are analysed, and normally both transaction logs (access logs) and search logs (logs of queries) are included in the term web logs. Transaction log analysis and *Search Log Analysis* on logs from the Web are thereby a part of Web analytics (Jansen, 2009b). But the concepts can also be viewed as a professional practice (Web analytics) and an academic research practice (transaction and search log analysis) with different purposes. Web analytics is used to pursue business goals on a website or to monitor the sites users,

and it is normally done using a software tool like Google analytics. Transaction and search log analysis are often done in more transparent ways and in closer encounter with the data in the log files, and usually in a retrospectively manner on historic log files.

The logs are secondary data from web servers. Jansen with colleagues states that a "transaction log is an electronic record of interactions that have occurred between a system and a user of that system" (Jansen et al., 2009, p. 2). With transaction log analysis it is possible to study both the aggregated behaviour at a macro level and the search patterns of individual users at a micro level of analysis (Jansen et al., 2009). Possible problems may be different format and content in the logs from the studied resources. And there may have been changes in the resources at some point and therefore there may be changes in the logs.

Transaction log analysis is rooted in behaviourism, but has a more open view than traditional behaviourism. In transaction log analysis the behaviour is studied without discounting the inner cognitive and affective aspects that accompanies the behaviour:

> "Research grounded in behaviourism always involves *somebody* doing *something* in a *situation*. Therefore, all derived research questions focus on *who* (actors), *what* (behaviors), *when* (temporal), *where* (contexts), and *why* (cognitive). The actors in a behaviourism paradigm are people at whatever level of aggregation (e.g. individuals, groups, organizations, communities, nationalities, societies, etc.) whose behavior is studied. Such research must focus on behaviors, all aspects of what the actors do. These behaviors have a temporal element, when and how long these behaviors occur. The behaviours occur within some context, which are all the environmental and situational features in which these behaviors are embedded. The cognitive aspect of these behaviors is the rational and affective processes internal to the actor executing the behaviors." (Jansen et al., 2009, p. 3)

In transaction log analysis behaviours are observed and classified as variables. Variables can be defined by their use or by their nature in the research. An example of use defined variables is independent or controlled variables. There are three types of nature defined variables: environmental, subject and behavioural. The environmental variables describe the contexts and the subject variables describes aspects of the subject studied, e.g. age or gender. The behavioural variables capture the observable activity of the subject, sometimes called trace data (Jansen et al., 2009). Jansen et al. lists six questions that must be addressed in every research project using trace data from log files. The questions address the issues of credibility, validity and reliability (Jansen et al., 2009, pp. 7-8):

1. Which data are analysed?
2. How is the data defined?
3. What is the population from which the researcher has drawn the data?
4. What is the context in which the researcher analysed the data?
5. What are the boundaries of the analysis?
6. What is the target of the interference?

In a log file it is often possible to single out specific search sessions and to follow the user through the session. It is probably not possible to track a user's different sessions, but it depends on IP-numbers and session-IDs. The data in the log files will be sufficient to explore the moves and tactics, the two most basic levels of the four levels of interaction (Bates, 1990; Jansen, 2009a) (see also Figure 5.2). Several interesting quantitative measures are relevant, e.g. bounce rate, depth and length of visits and the number of object looked at (Nicholas et al., 2006c). Log analysis is an evidence based research method, especially when log analysis is combined with user data, in some cases registered users (Nicholas et al., 2006c).

The main methodical issue is the limitations of log analysis. The method is built on a grounded theory approach, and is an unobtrusive method with several advantages (Jansen et al., 2009):

- Scale – large amount of data can be analysed
- Power – the large sample size of data makes it possible to do inference tests which can highlight statistically significant relationships
- Scope – all type of user-system interactions can be studied because all interactions are stored
- Location – logs can be collected in a natural environments and not in artificial settings
- Duration – log data can be collected over a long period of time

Major drawback with log data are (Jansen et al., 2009):

- As secondary data log files are not as versatile as primary data collected for the research questions in mind.
- The log data in not as rich as data collected by some other methods, and thereby cannot all kinds of research questions be answered.
- The fields recorded in the log might be very loosely linked to the concepts they are measuring
- Users may be aware that they are recorded and therefore change their behaviour/actions.

According to Jansen et al "all research methods suffer from some combination of abstraction, selection, reduction, context, and evolution problems that limit scalability and quality of results" (Jansen et al., 2009, p. 10). Maybe the most important aspects is the problem of abstraction, how to relate low level data to high level concepts. There are several drawbacks or challenges in log file analysis. If a proxy server is used may the complete sessions not be recorded in the log because the pages gets cashed in the proxy server and if a user revisits a previously visited page the cashed version is used and no request to the web site is made. The same phenomenon is present in normal browsers which cashes visited web pages for faster access when the page is revisited. Another challenge is the traffic caused by search engine spider, and any queries issued by the spiders, may be hard to distinguish from human users. It may also be difficult to determine the geographical location of the user and other demographics (Thelwall et al., 2005).

5.4.1 The studied log files

The log analysis is based on access log which covers three months, October, November and December 2010. The number of sessions was determined with a session timeout of 60 minutes in the software Web Log Storming[23]. In the analysis all major crawlers have been excluded. The sessions were analysed based on the navigation strategy used by the visitor to the site when arrived at the site.

All three of the resources have log files in the *Combined log*-format, one of the standard formats for web server logs (The Apache Software Foundation, 2012). The example (Figure 5.6) and explanations (Table 5.9) are based on the documentation on Apache HTTP Server Version 2.2 (The Apache Software Foundation, 2012):

```
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326 "http://www.example.com/start.html" "Mozilla/4.08 [en] (Win98; I ;Nav)"
```

Figure 5.6. An example of an access log file entry from (The Apache Software Foundation, 2012).

[23] http://www.weblogstorming.com/

Table 5.9. Explanation of the example in Figure 5.6 from (The Apache Software Foundation, 2012).

127.0.0.1	This is the IP address of the client (remote host) which made the request to the server. The IP address reported here is not necessarily the address of the machine at which the user is sitting. If a proxy server exists between the user and the server, this address will be the address of the proxy, rather than the originating machine.
–	The "hyphen" in the output indicates that the requested piece of information is not available.
frank	This is the userid of the person requesting the document as determined by HTTP authentication. If the document is not password protected, this part will be "-" just like the previous one.
[10/Oct/2000:13:55:36 -0700]	The time that the request was received. The format is [day/month/year:hour:minute:second zone] day = 2*digit month = 3*letter year = 4*digit hour = 2*digit minute = 2*digit second = 2*digit zone = (`+' | `-') 4*digit
"GET /apache_pb.gif HTTP/1.0"	The request line from the client is given in double quotes. The request line contains a great deal of useful information. First, the method used by the client is GET. Second, the client requested the resource/apache_pb.gif, and third, the client used the protocol HTTP/1.0.
200	This is the status code that the server sends back to the client. This information is very valuable, because it reveals whether the request resulted in a successful response (codes beginning in 2), a redirection (codes beginning in 3), an error caused by the client (codes beginning in 4), or an error in the server (codes beginning in 5). The full list of possible status codes can be found in the HTTP specification (RFC2616 section 10, http://www.w3.org/Protocols/rfc2616/rfc2616.txt).
2326	The last part indicates the size of the object returned to the client, not including the response headers. If no content was returned to the client, this value will be "-".
"http://www.example.com/start.html"	The "Referrer" (sic) HTTP request header. This gives the site that the client reports having been referred from. (This should be the page that links to or includes /apache_pb.gif).
"Mozilla/4.08 [en] (Win98; I ;Nav)"	The User-Agent HTTP request header. This is the identifying information that the client browser reports about itself.

Comments on the access log format: The IP-address and the system language indicate the country of origin or language of the user, and the IP-address also implies the usage of the resource in different countries (Gäde et al., 2010). The information about the user agent is used to sort human visitors from search engines spiders (bots) crawling the resource.

66.228.165.147 - - [01/Oct/2010:00:01:13 +0200] "GET /adl_pub/forfatter/e_forfatter/e_forfatter.xsql?ff_id=51&nnoc=adl_pub HTTP/1.1" 200 8087 "-" "Mozilla/5.0 (compatible; Yahoo! Slurp/3.0; http://help.yahoo.com/help/us/ysearch/slurp)"

Figure 5.7. An example of line in an ADL log.

In Figure 5.7 the domain adl.dk is not part of the recorded information as it implicit – the access log only saves the interactions at the particular web server. The user agent is the Yahoo! Search engine spider called Slurp. This interaction is not studied in the present research as it is excluded automatically by WLS because it is a spider, not a human.

h-96-214.scoutjet.com - - [26/Sep/2010:06:51:04 +0200] "GET /kid/VisWeilbach.do?kunstnerId=9352&wsektion=genealogi HTTP/1.1" 301 616 "-" "Mozilla/5.0

Figure 5.8. An example of line in a KID log.

In the example from KID (Figure 5.8) there is no referring information, just the information that the visitor used Mozilla Firefox 5. Probably the user arrived by direct navigation, a bookmark or by typing in the URL.

186.40.20.210 - - [01/Oct/2010:00:00:13 +0200] "GET /permalink/2006/poma/info/es/frontpage.htm HTTP/1.1" 200 23600 "http://abp-sil-colonia.blogspot.com/" "Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.1; Trident/4.0;

Figure 5.9. An example of line in a Poma log.

In Figure 5.9 the Poma-user arrived by a link from a Blogspot-blog and accessed the front page in Spanish. In Appendix 4 a whole sessions from ADL is presented as an example.

The log files for both ADL and KID contained some difficulties due to the technical structure of the content management system (CMS). Large parts of the content, especially the digitalized heritage objects, are only displayed after the system has generated a request. In ADL after .xsql? in the URL and after .do? in KID, as displayed below:

"GET /kid/VisWeilbach**.do?**kunstnerId=9352&wsektion=genealogi HTTP/1.1"

This means that the part of the URL before the question mark does not contain any information about which page that was requested. All the information about the page is after the question mark, in the example the artist with the id-number 9352 and within that page the section called "genealogi". In the logs from ADL and KID the extension ".html" was added to the end of the URL so it was treated as an html-page, and not a query in WLS.

5.4.2 Log pre-processing

The pre-processing of the log files includes data cleaning, user identification, session identification and path completion (Cooley et al., 1999). Embedded requests of files as parts of the web design were deleted because with every page request the system automatically requested for examples the style sheet just to get the graphical design correct. The embedded requests has nothing to do with the actions of the users, they are part of the technical design of the site. In the log files from ADL and KID the lines in which the requested URL ended with the one of file extensions css, ico, jpg, or gif were deleted, as they are non-essential, system-generated requests.

In the case of the log files for the Guaman Poma sub site, the logs were extracted from the whole set of log files for the web site of the Royal Library[24] with the grep command. The phrase/expression searched for was "2006/poma" based on the link path to the Guaman Poma sub-site with the URL:

http://www.kb.dk/permalink/2006/poma/info/en/frontpage.htm.

In the log files from ADL and KID some modifications had to be done because of the technical structure of the resource. The objects were stored in a dynamic database and thereby contained the requested URLs query elements (after the question mark in Figure 5.7 and Figure 5.8). As raw data the log files from ADL and KID could only be partially analysed in the log analysis software. The basic indicators could be obtained but not the paths of viewed pages because of the page requests were split into requested URL and query, and thereby were all objects of each type grouped together as one requested URL, e.g. VisWeilbach.do and not VisWeilbach.do?kunstnerId=9352 (Figure 5.8)

The following replacement where done within each log file, in the end of the GET-string:

- First " HTTP/1.1" was replaced with ".html HTTP/1.1" and " HTTP/1.0" was replaced with ".html HTTP/1.0" to add the html file extension to all lines so they could be treated as html pages in the log analyser and still contain the information in the queries (after the ? in the requested URL).
- Secondly the places where .html had been added to an existing .html where replaced with just one .html (".html.html" was replaced with ".html").
- Thirdly the question mark in the requested URLs was replaced with two dashes so the log analyser did not cut the URL into two halves automatically, the queries were possible included into the URL-analysis ("?" was replaced with "--").

In the log analysis software Web log storming there are a number of crucial settings. During the importing the log files all zero bandwidth hits were excluded. Those are the hits, requests, in the log file were no data was sent to the user/computer requesting the page, normally because the request was cancelled before the load was finished. All hits where the requests was automatically

[24] www.kb.dk

generated requests concerning elements in the graphical design were also excluded in the analysis. The excluded file types were css, jpg, ico and gif. The analysis was limited to human users only, other user agents like search engine spiders was automatically excluded based on the user agent information in the log files and lists of known spiders.

The internal queries were included (ADL and KID) in the URL during the analysis, otherwise it was impossible to see which objects that are requested and viewed. If the internal queries were left out from the URL-analysis just the stem of the requested URL was analysed, for example just www.kulturarv.dk/kid/VisVaerk.do without the specification of which artwork that are requested and not the whole www.kulturarv.dk/kid/VisVaerk.do?vaerkId=244002 where the id-number of the artwork is included. In the first case all requests to show artwork (VisVaerk) is treated as one URL, which means that it is impossible to measure how visited a single page is or which pages that are viewed.

No path completion, i.e. replacing missing entries in the log files due to the use of the back button in the browser or other cache issues, was done in the logs (Cooley et al., 1999). The focus on the aggregate session paths. In the present research design the use of for example the back button in the browser has no impact on the results as the previously visited page already is present in the log files. The distillation from the raw log files to the exported specific datasets was done in Web Log Storming[25] (WLS). Logs were loaded into WLS which placed the whole dataset in the RAM memory and thereby could the point of view in the data analysis be changed on the fly. The software lack export features, so datasets like the session was imported to a text editor by cut and paste. Key features in WLS are the possibilities to automatically exclude search engine spiders, create sub-datasets based on navigation strategy or referrer, and to display the hits in the logs per session. In Appendix 5 is a screen shot of WLS.

5.4.3 Human users versus search engine spiders

With the log analysis software Web Log Storming it is possible to divide the actors in the log files automatically into two groups based on the user agent information. Human users are separated from search engine spiders (often also called robots or bots)

Table 5.10. The distribution of sessions by human users, search engine spiders and unknown visitors in the log files.

	Human user	SE spider	All sessions	SE spider share
ADL	72519	70638	143157	**49%**
KID	28598	32711	61309	**53%**
Poma	51134	24542	75676	**32%**

[25] http://www.weblogstorming.com

The sessions cover all user activities in the three month analysis window. The search engine spider share of all the sessions was between 32% and 53% of the total number of sessions. Only the human user session was studied in the present research.

5.4.4 Session identification

The sessions are determined in length by a combination of IP-number and a session time out set to 60 minutes. After the number of minutes of inactivity the session is considered ended and a new session is started if the user returns. 60 minutes is chosen in the present study as a time out limit because of the nature of the material, mainly text. A time out between 5 and 120 minutes is often used during search log analysis (Jansen, 2009a), but the use of a search engine differs from a site with cultural heritage. Cultural heritage resources are content sites, a goal for information searching, whereas web search engines are tools used for seeking and the users uses them for transitions.

In the software Web Log Storming (WLS) the search engine navigation sessions and the link navigation sessions were sorted based on the strings in Appendix 6. The search engines were identified in the log files and the extraction was a straight forward process. The sessions based on the link navigation strategy were extracted by removing the direct navigation strategy sessions, which were a category in WLS, and the search engine navigation strategy sessions, as presented in Appendix 6.

5.4.5 Measuring path length, visited levels, and arrival level

To measure the length of the path, the number of pages visited, is straight forward. In each session every page view is represented by a cell in a spread sheet and each row is a session. The length of the path is equal to the number of cells with content.

To measure the levels visited in each session the page views in form of the requested URLs were transformed to general representations of the level they belong to, according to the site structure analysis. Every URL were replaced automatically by an N, I or O (see Figure 2.7a) by a search-and-replace procedure. The sessions were thus transformed from a list of URLs to an abstract string of letters, e.g. NIIIOIIO, based on the URL analysis in Appendix 8. In the next step the numbers of N, I, and O's were counted in each row to get the distribution of levels visited in each session.

The arrival level is equal to the level of the first viewed object in each session. Together with the path length and visited levels the arrival level forms the session paths that are analysed in Chapter 8.

5.4.6 Measuring the navigation strategy

An important indicator in the log files is the referrer. The referrer is the URL from the external web page in the first request in a session. It is the link in another site the user followed to get to the present site. If the referrer is a web search engine the search terms used in the search engine is normally included in the referrer, so here is a way to get a glimpse of the users' initial need or intention. Basically there are three types of referrers: web search engines, web sites, and direct (Levene, 2010), as discussed in Section 4.2.3. The referrers correspond to the navigation strategies on the web. Sessions starting with direct navigation are automatically sorted in WLS based on user-agent information.

The search engine sessions are sorted by a search for all the major referring search engines with a share of referring traffic over 0.05% (in Appendix 9). The link navigation sessions are all session minus the direct navigation sessions and the search engine navigation sessions. Because of this way of producing a link navigation session set (they cannot be derived directly in WLS) all session that is not defined as either direct or as a major search engine is seen as link navigation, including "unknown". Unknown are a category of hits in WLS where no user-agent information is delivered to the web server and thereby it is not possible to categorise the hits. It might be a human user with high privacy settings in their web browser or it might be spiders of some kind. The one page view unknown-sessions are balanced up with long sessions where the whole resources might be downloaded (e.g. HTTrack Website Copier[26]) or users arriving with some of the minor search engines. The unknown-sessions make up 7.4% of all sessions in ADL, 2.7% in KID, and 5.0% in Poma.

5.4.7 Analysis of queries in referring search engines

Among the data in the referring URL from a web search engine is the query that the user formulated and submitted in the search engine. Below is an example of a log entry with a query in the referring search engine URL. The query in Google is marked with bold and the spaces are replaced with plus-signs.

http://www.google.com/search?source=ig&hl=en&rlz=1W1GYWE_en&q=**guaman+poma+website**&aq=f&aqi=g1&aql=&oq=&gs_rfai=CM2F5MnWmTK60Daqytwf4qdUmAAAAqgQFT9BuxEw

In the log files the numerous queries are too many to analyse. A statistical sample of the queries from the referring search engines was gathered. In each resource every one hundredth query was extracted and analysed, e.g. query number 1, 101, 201, etc. The number of queries differed between the resources. The queries were analysed on two levels. First every query was classified as belonging to one or more of the three categories: informational, navigational, and transactional

[26] www.httrack.com

(Broder, 2002) depending on the search terms in the queries. Then was every informational query re-examined and classified according to the subcategories in Table 4.1. The subcategories were developed in an iterative process during the analysis of the queries, but they was also inspired by the categories used in the Getty study by Bates et al. (1993) and the Dublin Core metadata elements (The Dublin Core Metadata Initiative, 2012).

The categorisation of the search queries are always to some degree uncertain. It is impossible to determine if a query containing just the name of a author as search terms is an informational search to answer a topical information need or if the user known on beforehand that the query leads to the specific resource (a "hidden" navigational search).

5.4.8 Session paths

The session paths are based on the levels visited in the sessions. The sessions are categorised based on three indicators:

- Session length – one page visits vs. several pages visited
- Visited levels in the session
- Arrival level in the session

The sessions are categorised as one of the 15 path types in Table 5.10. In the paths longer than one page view the second and third arrow symbolically shows the levels visited beside the arrival level, for example in path N2 there can be several page views on the navigational level followed by page views on the informational level (the second arrow), and the page views on the navigational level.

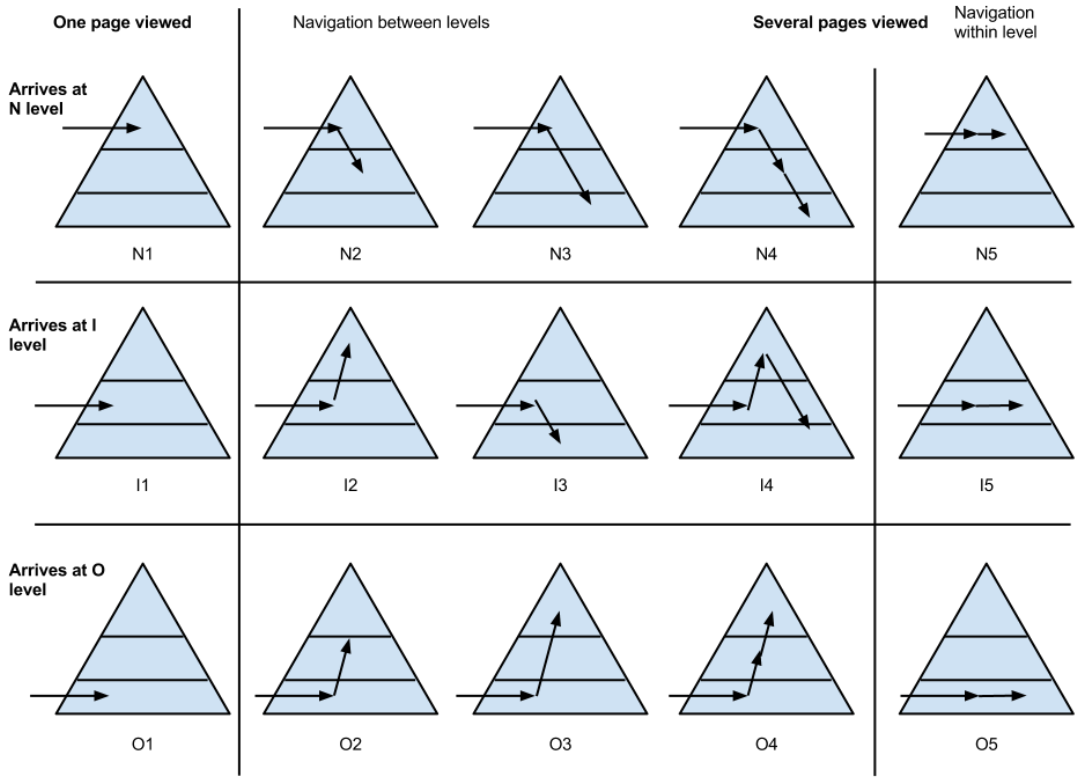


Figure 5.10. The 15 session path types in the three level resource model (based on Figure 2.7a).

The 15 session path types are used in the analysis of ADL and KID. For the analysis of the session paths in Poma a simplified version is used based on the two level version of the resource model without the informational level (see Figure 8.8) because of the type of content in Poma.

5.5 Web survey

5.5.1 Web survey as a method

The study combines access log data with data from a web survey. In Transaction Log Analysis (TLA) the behaviour of the user can be studied, i.e. *what* the user does. *Why* the user acts in the way he behaves is impossible to find out through analysis of logs. Therefore log analysis is often combined with a second method of data collection. The most common method is doing a survey to get answers about the why's and who's behind the interactions in the log files (Jansen, 2009b; Rainie & Jansen, 2009). Large-scale surveys can yield statistically generalized patterns of the users and their usage, but "this picture may remain at a coarse level and be contextually insensitive" (Savolainen, 2008, p. 77). For the RQ2 a web surveys have been used to get a richer description of the users. Web surveys were displayed on the front page of the resources (ADL and Poma) and as a pop-up (KID) studied for a limited time to obatin demographic data, and

information on intentions, goals and habits. The data collected through the surveys is primary data in form of quantitative, and some qualitative, survey answers.

One consideration in this study is information seeking in everyday life versus seeking in relation to work tasks (similarities and differences) in relationship to digitalized cultural heritage. Only a small part of the LIS theories deals with information seeking for leisure, and a large part of the usage of the cultural heritage collections could be with non-work intentions. In the log files all usage of the resources are captured; information searching for both leisure and work.

There are several benefits with self-administered questionnaire that collects data from respondents using the Web as the mode of data collection, especially the low resource use and the elimination of time and space boundaries (Tuten, 2010). And the participants tend to "feel more anonymous and therefore more honest in providing information to sensitive questions" (Tuten, 2010, p. 180). For a web survey posted on a web site the sample is non-probabilistic, a convenience sample. In non-probabilistic samples there is no way to "assess the potential magnitude of the bias, since there is generally no information on those who chose not to opt in" (Fricker, 2008, p. 199). These unrestricted, self-selected surveys have a crucial advantage that they can facilitate access to individuals that are hard to reach or identify, as the users of a specific web site (Fricker, 2008). All the users in the selected time gets the opportunity to answer the survey. The coverage error is all the frequent users of the resources which do not use the resource during the time of the survey. Normally sampling error and measurement error are important, but they are not applicable on the present surveys as the participants volunteered and the samples are not statistically representative. On the contrary, it might be misleading to report sampling error for opt-in samples as the measurement signals representativeness (Fricker, 2008).

5.5.2 Survey questions

The questions are discussed below. The questions in all three languages are described together with the answer are in Appendix 7. The questions cover different areas on the user and the visit (as discussed in Chapter 4). For the majority of the questions there were fixed choice answers. Self-reporting might distort the data by people's perceptions of their own behaviour and skills (Holt, 2012), but the advantages of using a web survey is greater than the disadvantages which are present in all usage of interviews or surveys. Some of the questions in the questionnaire are inspired by Mette Skov in a study of information seeking behaviour in a Danish museum web site (Skov, 2009).

To capture the navigation strategy used the participants in the survey answered the question "How did you reach this site?" and it was possible to choose one of five different answers instead of just the three navigation strategies discussed in Section 4.2.3. Direct navigation was split into navigation through a bookmark and navigation by typing in the URL. Search engine navigation was split into two types of searches: informational and navigational (Broder, 2002). The splitting of Levene's three navigation strategies (2010) was done to capture some finer aspects of respectively strategy.

The intention with the visit is important because not all seeking behaviour has as goal to find specific information. Exploratory search has risen as an important concept that complements the look-up or known-item search (Marchionini, 2006a). The users were asked why they visited the resource and they were given the choices: *exploring*, *learning* and *look up fact*, based on the concept of exploratory search, together with *curious* and *other*. In this particular case it was possible to choose more than one answer, due to the fact that the intentions may be nested.

The task context measured in the survey is a combination of work task, information seeking task and information searching task together with the non-work task concept of coincidence. In IS&R-research the focus has, historically seen, been on the information behaviour in work contexts (Case, 2007). During the 1990s the learning context got attention (e.g. Limberg et al., 2002), and in the 2000s the leisure context was more broadly introduced (Stebbins, 2007), as well as everyday life as a context (e.g. Savolainen, 2008). The possible answers to the survey question "In what context do you visit the web site?" are: *School or study visit*; *Hobby or leisure*; *Work*; *By coincidence*; or *Other (please specify).*

The users' familiarity with the resource might influence their interactions. Regular visitors are probably less prone to bounce (leave after one page view).

The level of web search skills is based on three questions in the survey where the participants rated their knowledge and skills around information seeking on the Internet. The respondents had to rate themselves on a four point scale, ranging from excellent to bad. The three questions:

How do you rate your knowledge about the Internet?

How do you rate your skills in using Internet and Web technologies, e.g. using web browsers, web search engines and other web tools?

How do you rate your ability to evaluate information on the Web with regard to its relevance, quality and credibility?

The questions are based on the web search competencies discussed in Chapter 2.

Five demographic attributes of the respondents were asked for in the survey: *Age*; *Gender*; *Number of years in school*; *Country of residence*; and *Present position* (open ended).

The possible answers to the question about the respondents country of residence are the continents, except in Europe where Denmark and Scandinavia are alternatives on the own due to different reasons; Denmark because the Danes are the overall largest user group of the Danish cultural heritage resources; and the other Scandinavian countries because of their closely related languages and shared older history. The answer to the question gives a indication of how the usage is distributed over the world, but no evidence about the nationality or background of the users. For example all the users in Asia could be Danes living abroad. The distribution of countries can be compared to both the language version of the survey chosen and with the nature of the cultural heritage resource, e.g. Poma will for instance probably attract other users than ADL due to their different languages of the full text.

5.5.3 Deployment of the survey

The survey was launched in three language versions; Danish and English in ADL and KID, and English and Spanish in Poma. It was planned that invitations to the surveys should be pop-up windows appearing directly when the user arrived at the resource. In reality the pop-up invitations were only deployed in KID. In ADL and Poma the invitation was published on the top page due to unexpected technical reasons. The invitation contained two short texts in the corresponding languages with an invitation and a link to the survey. The survey consisted of eleven questions (see Appendix 7) and was online for 20 days in KID and 23 days in ADL and Poma, in January and February 2012. The lack of a pop-up invitation had an impact on the number of responses. In the studied log files just 16% of the users visit a page on the navigational level where the top page is. The number in Poma is 42%, but still the majority of the users do not even visit the top, navigational level. In KID 256 participants answered at least 10 of the 11 questions, but in ADL the number was 56 and in Poma just 44 participated in the survey.

The sampling is "volunteer, accidental, convenience" as Black classifies it (Black, 1999, pp. 118, 125). The sampling technique is inexpensive, but can be highly unrepresentative. If the users are not known, there is not many methods of sampling to use. In fact it might be the only possible way to reach unregistered users of a website. Due to the method of sampling it is impossible to talk about representativeness. All respondents are users of the studied resources, but which users that did take part in the surveys are unknown. Due to the volunteer sampling the sample covers an unknown part of the population of users of each cultural heritage resource, and it is impossible to say how representative the sample is. The answers to some of the questions were compared with data from the log files and the degree of representativeness is discussed in the conclusion.

5.6 Mixing methods

5.6.1 Data types

The different data collection methods generate different types of data that can be divided into several types. The first distinction is between primary and secondary sources, and thereby primary and secondary data. What this distinction between the two types is based on differ. One distinction is depending on the closeness in time between the event and the recording. Primary data is recorded directly or close in time, and secondary sources interpret or record primary data (Walliman, 2011, p. 69). Another distinction is based on the reason that data collected first hand for the research by the researcher, or by an assistant, is primary data; and secondary data is collected by others for other reasons, e.g. census data (Frankfort-Nachmias & Nachmias, 2000, p. 276). The two different distinctions pin-point two different aspects on the data. The first is about the relation between the event and the recording, if the recordings of the event are

simultaneous, contemporary or of a late date. The second one is about the control over the data collection and the adoption of the data to the research questions.

The data in the present study have different conditions. The data in the log files is recorded instantly when the actions occur (close in time), but the data is not primary collected for this study. The data is collected automatically by the system to gather usage statistics and for tracking possible errors. Depending on view, the log data might be seen as either primary or secondary data. If the logs were customized for the research then they would be seen as primary, like for example Gäde et al (2010). The web survey data is also collected close in time, when the participants visit the resource, and the questions are mainly about the current visit. The questions in the survey are designed for the present research and to some extent based on preliminary findings in the logs. Due to the closeness in time between the survey and the event (the session) the data is seen as primary in both distinctions. For both the site structure analysis and the findability analysis data is collected directly on the web (primary data), and because of the invariant character of the resources studied there have been no larger changes over time.

5.6.2 Combing different methods

In the research design the three methods are employed as a multistage rocket (Figure 5.3). The first stage is the site structure analysis which classifies the URLs into one of the three levels. The second stage is the log analysis were the sessions are identified and studied based on average number of page view, referring navigation strategy and arrival level, together with the findability analysis. The third stage is the web survey asking questions about the context of visit to the site and is parallel to the other methods. This multistage rocket design connects the content and structure of the site with usage data in the log files and the users and their context in the survey.

5.6.3 Time overlap in data collection

One drawback of the present design is that the time periods are not overlapping, the logs and the survey covers different periods. But on the other hand both datasets contains real users with their own intentions and tasks which has led them to the site. And if the survey was distributed during the same time as logs were collected, the pop-up survey might have interfered with the usage of the site. Log analysis was chosen because through the method it is possible to study real users in everyday life situations, unobtrusively. If the survey and logs was covering the same period of time the respondents would be among the sessions in the logs, but it would still be impossible to say how representative they were because it is not possible to count the number of users instead of the number of sessions, or the sessions of the respondents. The only possibility to connect the respondents and their answers with interactions in the logs would be to identify the survey participants in the logs and then just study their activities.

5.6.4 Potential patterns

In quantitative research reliability and validity are central concepts. There are four types of validity in pure quantitative research: face validity; content validity; criterion validity; and, construct validity (Neuman, 2006). Kvale has a more qualitative approach to the concept of validity when he discusses interviewing:

> "Validation rests on the quality of the researcher's craftsmanship throughout an investigation, continually checking, questioning and theoretically interpreting the findings." (Kvale, 2008, p. 123)

As stated in the quote above, Kvale sees validity as having three dimensions. The first is the checking for invalidity; a continuous critical look on the analysis and findings to attempt to falsify it. The second is validating by questioning if the method is the right one for the purpose of the study; different types of research questions lead to different methods. Together with reliability and representativeness, validity forms the base which generalisations rest upon. In a general sense reliability is to what extent the observations indeed reflects the phenomena or variables investigated. Despite the fact that the results sometimes cannot be seen as statistically representative, there are other ways of talk about generalisation. When discussing interviews as a research method Kvale presents three types of generalisations: naturalistic generalisation, statistic generalisation, and analytic generalisation. The first type is based on personal experience and tacit knowledge. Statistic generalisations are formal and explicit, and are based on random samples from the population. This type of generalisation can be drawn for small samples as long the participants are randomly selected, and not e.g. volunteers. The third type, the analytic generalisation, is based on a reasoned judgement about the extent to which the findings of one study can be applied in other settings or situations. It is based on the analysis of similarities and differences between the two situations. Both the researcher and the reader can do analytic generalisations, the researcher has to argue for the generalisations in the text and the readers do them on the basis of the context descriptions (Kvale & Brinkmann, 2009, pp. 281-284).

In quantitative research reliability and validity are central concepts, but in the present study neither the reliability nor the validity leads to representative generalisations of the findings as the gathered data is not collected in a statistically representative manner. The log analysis is a total analysis of the whole logs, the sample in the web surveys are volunteer samples, and the findability analysis is done on typical objects from the different resources and levels. As briefly discussed in Fransson (2012) is it probably better to talk about the typicality of the findings in the datasets. Typical patterns might be discovered through all the four methods in the research design, and any patterns will be seen as pictures of the resources, the findability, the usage and the users. Together these findings can highlight possible typical connections between the levels of analysis. One example of typical pattern is the navigation strategy the visitors uses to get to the resources. A small group arrives by direct navigation, by bookmarks in the browser or by typing in the URL, in all of the studied resources. At the same time a very large group navigates to the resources by a web search engine, most often Google. The use of navigation strategies is a pattern, and there

is no reason to believe that other cultural heritage resources, similar or smaller, are not navigated to in the same manner. Generally there are probably more common patterns between the cultural heritage resources than there are different and unique patterns in the individual resources.

5.7 Chapter summary

The central question in the chapter is: How can an appropriate research design be shaped to collect both usage data and findability data from cultural heritage resources? A mixed methods research design was chosen to capture data, both usage and findability, from different perspectives. The design was based on the URI model (Figure 2.11) and the goal with the model was that each method should focus on specific dimensions, e.g. the log analysis is primarily focused on the actions of the users. The mixed methods design was both parallel and sequential. The survey was deployed independently of the other methods, i.e. parallel to them. The site structure analysis preceded the log analysis and the findability analysis as the later methods were based on the results of the site structure analysis.

6 The findings of the site structure analysis

6.1 The site structure of the resources

The site structure analysis is presented first because the results are used in both log analysis and the findability analysis (see Figure 5.3). In the analysis the content in the resources is divided into levels. Every function and object is classified as belonging to one of the levels of the resource, in the case of Poma into two levels. The analysis of sessions in the log data is based in the site structure analysis. All page (object) views are transformed into level and are seen as page views on a specific level or transitions between levels, rather than views of specific objects.

The objects were in the site structure analysis classified as belonging to one of the levels: navigational (N), informational (I), or object (O) based on the object-resource-framework (seeFigure 2.7a) and their URL. Appendix 8 contains both the URL-analysis and the site structure analysis. All the listed objects represent a class of objects of the same type, and the objects in the lists cover all types of objects in each resource. All listed objects are classified as N, I or O.

The analysis was guided by the framework in Chapter 2, but it was also an interpretative process. For example, in KID the information about the artist and the museums was classified as belonging to the upper, more general informational level and not the lower (cultural heritage) object level because of the aim of the study – to see if the users actually reach the digitalised cultural objects, in KID the artwork information. But it could be argued that the artist or museum information also is cultural heritage objects and not just a supplement to the artworks.

The result of the site structure analysis is displayed in Figure 6.1, Figure 6.2 and Figure 6.3 below. In the figures the different functions or different kinds of objects are placed at the level they are classified at. See Appendix 8 for the categorisation for each type of object.

In ADL as well as in the other resources the general information together with navigation and search features were classified as belonging to the navigational level (N). The literary works in full text were classified as belonging to the object level in the bottom. These objects are the core of the resource as it is created to mediate them. The texts were available in at least one of the three formats: text, facsimile and unformatted downloadable text. The text-versions are presented as normal text in html and are searchable with the search function of the browser. The facsimile is an image of the digitalised printed page and is presented as a picture. The last type, the downloadable text is the whole literary work as a single text file (ASCII). The first two formats present one book page per object. The pages with author information, biography, list of titles and bibliography, as well as the information about literary periods were categorised as informational (I). They were neither general like the objects at the navigational lever nor specific as the literary texts on the object level (O).

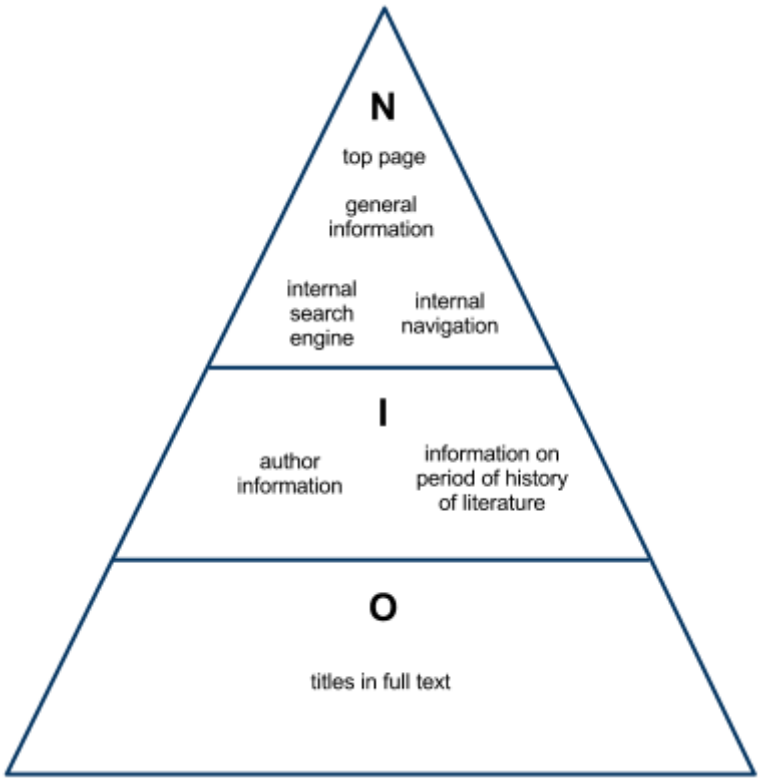


Figure 6.1. The site structure of ADL in the resource-model.

The structure of KID is similar to the structure of ADL. In KID the object level contains the information on individual artworks with title, artist, holding, some metadata and sometimes a small picture of the artwork. At the informational level the amount of artist information varies greatly. In KID the artist encyclopaedia Weilbachs Kunstnerleksikon is integrated. In KID there are about 24,700 artists represented, but only 8,000 are mentioned in Weilbachs Kunstnerleksikon and has thereby more artist information than the two thirds not mentioned in Weilbachs. Some of the artists are just presented by their names and have no artworks listed.

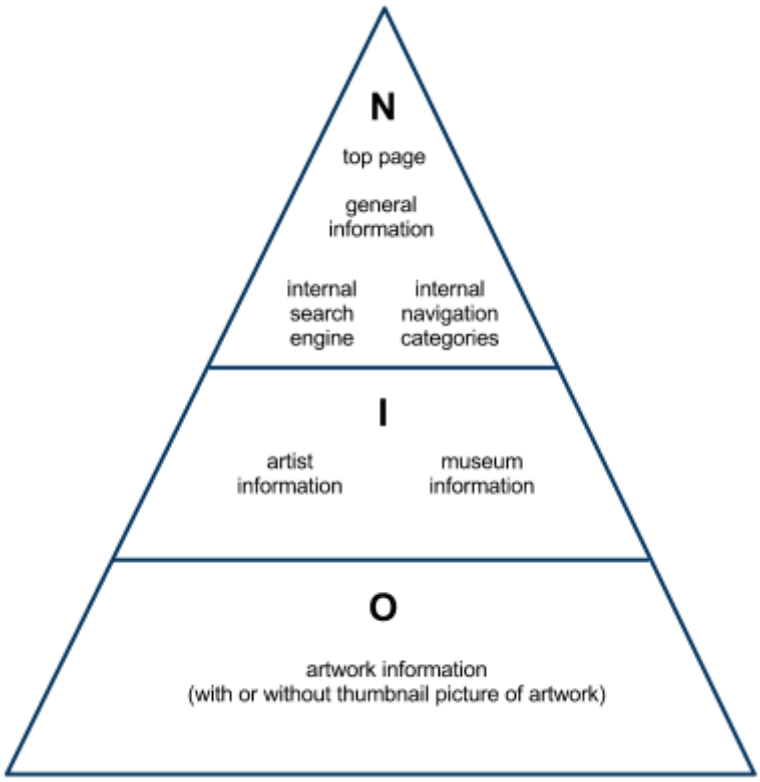


Figure 6.2. The site structure of KID in the resource-model.

The Guaman Poma sub-site is completely focused on the digitalized manuscript of the Inca Chronicle. Therefore is the resource divided into just two levels; the general information is only about the Poma Chronicle and is not possible to divide into two levels – a general and a more specific. Here N is used for the level with the information about the resource, and O is used for the level with the digitalized pages.

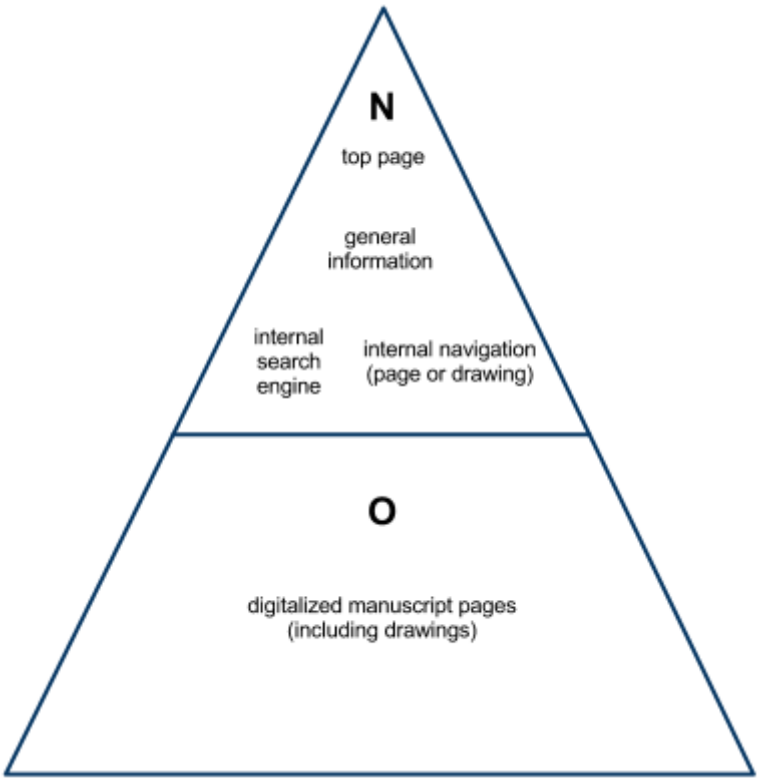


Figure 6.3. The site structure of Poma in the resource-model.

6.2 Chapter summary

The chapter presents the results of the analysis of the three cultural heritage resources. The function of the structure and content was studied and all three resources were found to have a navigational level and an object level in the site structure analysis. ADL and KID also have a level in-between, an informational level with information about the objects in the object level and about their creators, authors and artists. All types of URLs were categorised as navigational (N), informational (I) or object (O). The site structure analysis is used both to describe the resources and in the log analysis (Chapter 8) and in the findability analysis (Chapter 7).

It is also a tool for describing the resources as shown in Figure 6.1, Figure 6.2 and Figure 6.3. The analyses are interpretative and the results might to some extent depend on the aim of the study, but that is a strength of the site structure analysis as it serves as a support method, for both the findability analysis (Chapter 7) and the log analysis (Chapter 8).

7 How findable are the resources?

The chapter focuses on the first research question: *How findable is the heritage resources and their objects?* The findability is studied by an evaluation of central aspects. The evaluation tries to create an illustration of the level of findability, both external and internal, on the different levels in each resource. The theoretical background is in Chapter 3 and the methodology discussed in Chapter 5. How the theoretical concepts are turned into evaluable variables and how they are applied in the present study, is in Chapter 5.

The full results of the evaluations of the selected objects can be found in Appendices 10-14, and the results are summarised in Section 7.2. This is followed by calculations of the external findability as well as the internal findability. The findability scores are used to discuss the findability of the objects on the different levels, and which impact they may have on the usage. In the summary the two sub-research questions focusing on the empirical aspects of the findability analysis are answered.

7.1 Studied objects

Typical objects on each level were chosen to represent all objects on the level. Both popular and easily accessible objects as well as non-popular and harder to access-objects were selected. On the navigational level all central objects, i.e. top page and interal search, were included. On the lower levels representative objects were chosen to illustrate the indicators, and to illustrate the method. Between 14 and 15 objects are studied in each resource. The objects are distributed on the levels according to Figure 7.1.

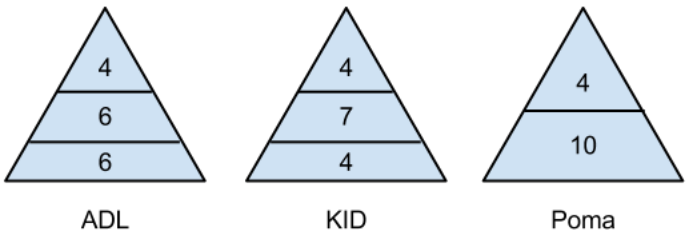


Figure 7.1. The number of objects on each level studied in the findability analysis.

In Table 7.1, Table 7.2 and Table 7.3 the evaluated objects are listed. Each object has an ID for easy referring and to avoid mix up.

Table 7.1. The studied ADL objects.

Id	Level	Title	URL
A-N1	Navigation	Top page (first)	http://adl.dk/adl_pub/forside/cv/forside.xsql?nnoc=adl_pub
A-N2	Navigation	Author list	http://adl.dk/adl_pub/forfatter/forfatter_menu.xsql?nnoc=adl_pub
A-N3	Navigation	Search, page one, introduction	http://adl.dk/adl_pub/soeg/cv/search_menu.xsql?nnoc=adl_pub
A-N4	Navigation	Search, page two, free text in texts and author biographies	http://adl.dk/adl_pub/soeg/cv/fritekst/fritekst_soegning.xsql?nnoc=adl_pub
A-I1	Information	Author first page: H.C. Andersen	http://adl.dk/adl_pub/forfatter/e_forfatter/e_forfatter.xsql?ff_id=22&nnoc=adl_pub
A-I2	Information	Author title list: H.C. Andersen: D	http://adl.dk/adl_pub/vaerker/cv/ff_vaerker_menu.xsql?ff_id=22%20&bogstav=D&nnoc=adl_pub
A-I3	Information	Versions of the Little Mermaid	http://adl.dk/adl_pub/vaerker/cv/e_vaerk/e_vaerk.xsql?ff_id=22%20&id=2247&hist=fmD&nnoc=adl_pub
A-I5	Information	Author first page: Jacob Worm	http://adl.dk/adl_pub/forfatter/e_forfatter/e_forfatter.xsql?ff_id=44&nnoc=adl_pub
A-I6	Information	Author title list: Jacob Worm	http://adl.dk/adl_pub/vaerker/cv/ff_vaerker_menu.xsql?ff_id=44&bogstav=&nnoc=adl_pub
A-I7	Information	Versions of Annike Bi	http://adl.dk/adl_pub/vaerker/cv/e_vaerk/e_vaerk.xsql?ff_id=44&id=5073&hist=fm&nnoc=adl_pub
A-O1	Object	The Little Mermaid, facsimile, p.1.	http://adl.dk/adl_pub/pg/cv/ShowPgImg.xsql?p_udg_id=93&p_sidenr=87&hist=&nnoc=adl_pub
A-O2	Object	The Little Mermaid, facsimile, p.3.	http://adl.dk/adl_pub/pg/cv/ShowPgImg.xsql?nnoc=adl_pub&p_udg_id=93&p_sidenr=89
A-O3	Object	The Little Mermaid, text, p.1.	http://adl.dk/adl_pub/pg/cv/ShowPgText.xsql?p_udg_id=93&p_sidenr=87&hist=&nnoc=adl_pub
A-O4	Object	The Little Mermaid, text, p.3.	http://adl.dk/adl_pub/pg/cv/ShowPgText.xsql?nnoc=adl_pub&p_udg_id=93&p_sidenr=89
A-O5	Object	The Little Mermaid, downloadable text	http://adl.dk/adl_pub/pg/cv/AsciiPgVaerk2.xsql?nnoc=adl_pub&p_udg_id=93&p_vaerk_id=2247
A-O6	Object	Annike Bi, facsimile	http://adl.dk/adl_pub/pg/cv/ShowPgImg.xsql?p_udg_id=175&p_sidenr=30&hist=fm&nnoc=adl_pub

On the object level all three versions of the Little mermaid was studied to examine the differences between facsimile, text and downloadable text. In addition to page 1, the start of the text, page 3 was also studied, expect in the downloadable version where the whole story is included in one unformatted text page.

Table 7.2. The studied KID objects.

Id	Level	Title	URL
K-N1	Navigation	Top page	https://www.kulturarv.dk/kid/Forside.do
K-N2	Navigation	Advanced search for artists	https://www.kulturarv.dk/kid/SoegKunstner.do
K-N3	Navigation	About Kunstindeks Danmark	https://www.kulturarv.dk/kid/OmKID.do
K-N4	Navigation	Museums in Kunstindeks Danmark [A]	https://www.kulturarv.dk/kid/SoegMuseumsoversigt.do
K-I1	Information	Artist: Karen Abell	https://www.kulturarv.dk/kid/VisKunstner.do?kunstnerId=8135
K-I2	Information	List of artwork by Karen Abell	https://www.kulturarv.dk/kid/SoegKunstnerVaerker.do?kunstnerId=8135
K-I3	Information	Information from Weilbachs Kunstnerleksikon: Karen Abell (all) (categories: genealogy, exhibitions, travels, education, occupations, biography, artworks, scholarships, literature, all)	https://www.kulturarv.dk/kid/VisWeilbach.do?kunstnerId=8135&wsektion=alle
K-I4	Information	AROS – Aarhus Kunstmuseum	https://www.kulturarv.dk/kid/VisMuseum.do?museumId=528
K-I5	Information	Artist: F.M.E. Fabritius De Tengangel	https://www.kulturarv.dk/kid/VisKunstner.do?kunstnerId=6574
K-I6	Information	Information from Weilbachs Kunstnerleksikon: F.M.E. Fabritius De Tengangel (all) (categories: genealogy, exhibitions, travels, education, occupations, biography, artworks, scholarships, literature, all)	https://www.kulturarv.dk/kid/VisWeilbach.do?kunstnerId=6574&wsektion=alle
K-I7	Information	Artist: Berenice Abbott	https://www.kulturarv.dk/kid/VisKunstner.do?kunstnerId=19742
K-O1	Object	Artwork 1 by Karen Abell (“Uden titel”)	https://www.kulturarv.dk/kid/VisVaerk.do?vaerkId=450177
K-O2	Object	Artwork 4 by Geskel Saloman (“Portræt af Moses og Hanne Ruben“)	https://www.kulturarv.dk/kid/VisVaerk.do?vaerkId=501506
K-O3	Object	Artwork 14 by Fabritius de Tengnagel, F.M.E. (”Vinterlandskab fra Langeland”)	https://www.kulturarv.dk/kid/VisVaerk.do?vaerkId=94722
K-O4	Object	André Maurois by Berenice Abbott	https://www.kulturarv.dk/kid/VisVaerk.do?vaerkId=413089

In Poma half of the studied objects are within an English framework (menus, etc.) and the other half within a Spanish version. The content is the same for both language versions, but there may be differences in how parallel objects (English/Spanish versions) are indexed by the web search engines or how many inlinks they have. In the evaluation both versions of each object is included, the odd numbered objects are English (i.e. Poma-N1) and the even-numbered is Spanish (i.e. Poma-N2). The only difference in the URL:s are the folders in the file structure, EN and ES, otherwise they are identical.

At the navigational level the top page was chosen together with the digital resources-page, a typical page with general content. Five objects were chosen at the object level: title page, first page, page 2, page 79 and page 80. The title page, the first page and page 79 are easily accessed in the table of contests (the left side menu) and are natural starting points for users. Pages 2 and 80 are not possible to access directly in the left side menu (the TOC) and are the second page in their respective chapter, and thus they are of the same kind as the majority of the objects in the resource.

Table 7.3. The studied Poma objects.

Id	Level	Title	URL
P-N1	Navigation	Top page (front page) English	http://www.kb.dk/permalink/2006/poma/info/en/frontpage.htm
P-N2	Navigation	Top page (frontage) Spanish	http://www.kb.dk/permalink/2006/poma/info/es/frontpage.htm/
P-N3	Navigation	Digital resources - English	http://www.kb.dk/permalink/2006/poma/info/en/docs/index.htm
P-N4	Navigation	Digital resources – Spanish (Recursos digitales)	http://www.kb.dk/permalink/2006/poma/info/es/docs/index.htm
P-O1	Object	Title page (Drawing 0 [1]) - English	http://www.kb.dk/permalink/2006/poma/titlepage/en/text/
P-O2	Object	Title page (Drawing 0 [1]) - Spanish	http://www.kb.dk/permalink/2006/poma/titlepage/es/text/
P-O3	Object	Page 1 - English	http://www.kb.dk/permalink/2006/poma/1/en/text/
P-O4	Object	Page 1 - Spanish	http://www.kb.dk/permalink/2006/poma/1/es/text/
P-O5	Object	Page 2 (Drawing 2) - English	http://www.kb.dk/permalink/2006/poma/2/en/text/
P-O6	Object	Page 2 (Drawing 2) - Spanish	http://www.kb.dk/permalink/2006/poma/2/es/text/
P-O7	Object	Page 79 (Drawing 23) (first page of Chapter 6) - English	http://www.kb.dk/permalink/2006/poma/79/en/text/
P-O8	Object	Page 79 (Drawing 23) (first page of Chapter 6) - Spanish	http://www.kb.dk/permalink/2006/poma/79/es/text/
P-O9	Object	Page 80 - English	http://www.kb.dk/permalink/2006/poma/80/en/text/
P-O10	Object	Page 80 - Spanish	http://www.kb.dk/permalink/2006/poma/80/es/text/

7.2 Evaluation of findability aspects

In Figure 7.2 the findability aspects are placed in the resource model and the object model are reproduced for ease of reading (originaly in Figure 3.5).

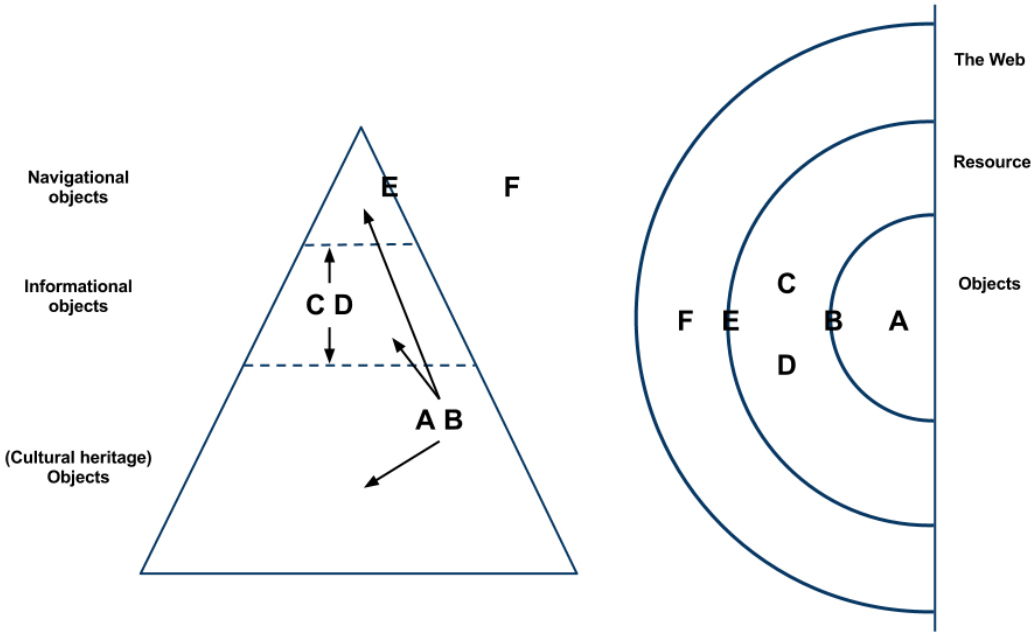


Figure 7.2. Findability aspects in the resource model and in the object model. The same model as in Figure 3.5, reproduced for for ease of reading. (A. Object attributes. B. Accessibility. C. Internal navigation. D. Internal search. E. Reachability. F. Web prestige.)

All studied objects were reachable by internal links, in all three resources. And on all three resources the internal search engines was covering parts of the content with a focus on the informational and object levels. It was not possible to find the objects on the navigational level in any of the internal search engines. All studied objects were reachable from the web through unique and stable links. Allmost all links were long and were therefore hard to type in for direct URL navigation.

Table 7.4. Summary of ADL findability evaluations.

Id	Object Attributes	Accessibility	Internal navigation	Internal search	Reachability	Web Prestige	External findability	Internal findability	Total findability
A-N1	1	2	1	0	1	3	**7**	**4**	**8**
A-N2	2	2	1	0	1	3	**8**	**5**	**9**
A-N3	2	2	1	0	1	3	**8**	**5**	**9**
A-N4	2	2	1	0	1	3	**8**	**5**	**9**
A-I1	2	2	1	1	1	3	**8**	**6**	**10**
A-I2	2	2	1	1	1	2	**7**	**6**	**9**
A-I3	2	2	1	1	1	0	**5**	**6**	**7**
A-I4	2	2	1	1	1	2	**7**	**6**	**9**
A-I5	2	2	1	1	1	0	**5**	**6**	**7**
A-I6	1	2	1	1	1	0	**4**	**5**	**6**
A-O1	1	2	1	0	1	2	**6**	**4**	**7**
A-O2	1	2	1	0	1	0	**4**	**4**	**5**
A-O3	3	2	1	1	1	2	**8**	**7**	**10**
A-O4	3	2	1	1	1	0	**6**	**7**	**8**
A-O5	1	3	1	0	1	0	**5**	**5**	**6**
A-O6	1	2	1	0	1	0	**4**	**5**	**5**

In ADL there are two kinds of internal search, search for titles and search in text pages and author portraits. The "text page"-search does not search in facsimile texts or the unformatted text for download, they are not indexed, so a large share of the full text is not findable through internal search. The internal link navigation is straightforward, because links to all full text are found on the author pages. The texts often contains several pages. The first page in each work is the most findable, as it is normal to link to the beginning of a work. The only object with the highest accessibility score is ADL-O5, due to the fact that it does not suffer from any WCAG-compliance problems according to AChecker.

Table 7.5. Summary of KID findability evaluations.

Id	Object Attributes	Accessibility	Internal navigation	Internal search	Reachability	Web Prestige	External findability	Internal findability	Total findability
K-N1	2	2	1	0	1	0	**5**	**5**	**6**
K-N2	2	2	1	0	1	0	**5**	**5**	**6**
K-N3	2	2	1	0	1	0	**5**	**5**	**6**
K-N4	2	2	1	0	1	0	**5**	**5**	**6**
K-I1	2	2	1	1	1	0	**5**	**6**	**7**
K-I2	1	2	1	1	1	0	**4**	**5**	**6**
K-I3	3	2	1	0	1	0	**6**	**6**	**7**
K-I4	1	2	1	0	1	0	**4**	**4**	**5**
K-I5	2	2	1	1	1	0	**5**	**6**	**7**
K-I6	3	2	1	0	1	0	**6**	**6**	**7**
K-I7	1	2	1	1	1	0	**4**	**5**	**6**
K-O1	2	2	1	1	1	0	**5**	**6**	**7**
K-O2	2	2	1	1	1	0	**5**	**6**	**7**
K-O3	2	2	1	1	1	0	**5**	**6**	**7**
K-O4	2	2	1	1	1	0	**5**	**6**	**7**

In KID there was a general mix of languages in the interface. The cultural heritage objects contain information in Danish, but when English is chosen the menus are in English so the SAPs are in two language. It is the total amount of SAPs that is counted for, regardsless of languages it is only counted once. In many cases the menus and other fixed content in the template accounts for at least half of the number of SAPs. KID was indexed by Google, but the objects were not ranked. No PageRank was calculated for the objects, probably because the secure version of the transfer protocol HTTP is used (Hypertext Transfer Protocol Secure, HTTPS). The consequence is that the objects from KID are low ranked in comparison with other web pages.

Table 7.6. Summary of Poma findability evaluations.

Id	Object Attributes	Accessibility	Internal navigation	Internal search	Reachability	Web Prestige	External findability	Internal findability	Total findability
P-N1	2	2	1	0	1	2	7	5	8
P-N2	2	2	1	0	1	3	8	5	9
P-N3	2	2	1	0	1	2	7	5	8
P-N4	2	2	1	0	1	2	7	5	8
P-O1	3	2	1	1	1	0	6	7	8
P-O2	3	2	1	1	1	2	8	7	10
P-O3	3	2	1	1	1	2	8	7	10
P-O4	3	2	1	1	1	2	8	7	10
P-O5	2	2	1	1	1	0	5	6	7
P-O6	2	2	1	1	1	0	5	6	7
P-O7	2	2	1	1	1	2	7	6	9
P-O8	2	2	1	1	1	0	5	6	7
P-O9	3	2	1	1	1	0	5	7	7
P-O10	3	2	1	1	1	0	5	7	7

In Poma the internal search engine only searches: *Transcript, Normalized Quecha, Commentary and other notes*, and *The entire text*. The text in the table of contents (TOC) is not searchable, which is the most important navigational tool. Some of the studied objects in Poma were not ranked in Google, but indexed. These consisted of two kinds of material: duplicates were the one of the language versions was ranked but not the other one; and objects a step or two from the start of a chapter, in practice objects further down in the link structure.

7.3 External and internal findability

The concept of external findability tries to measure how easy an object or a whole resource is to be found from the surrounding web. External findability is the sum of the aspects important for navigation from the outside (see Table 5.8).

A general tendency is that the lower score in the range of external findability drops at the Information (I) and Object (O) levels compared with the Navigation (N) level in both ADL and Poma. In KID the pattern is different, both the lowest and highest score are at the Information (I) level. The difference between the resources is due to the impact of web prestige. Objects in top of the resource often have a similar PageRank-value based on the closeness to the top page, which gets most inlinks, and the nature of PageRank link analysis – some of the PageRank is inherited by internal linking. In the case of KID none of the objects have a PageRank-value and thereby the external findability is determined by the other variables. Within the present external findability evaluation framework and the studied resources it is the web prestige aspect that differs the most and it determines the level of external findability in many cases. But at the same time the web prestige determines the probability that a user finds the object among all other objects.

The internal findability was to a large degree dependent on the number of SAPs, in combination with internal search for which in all three resources only the objects on the object and informational levels were indexed. All objects were accessible and findable by internal link navigation.

The range of the normalized findability scores in the figures below is based on the lowest score and the highest of the objects on each level. As shown in Figure 7.1 the number of studied objects on each levels differs between four and ten. The ranges of the findability scores differ between the external (0-10 points) and the internal (0-8 points) depending on the aspects included. When the findability scores are normalized and expressed as percentages of the maximum score within the framework the differences between the levels becomes clear (see Figure 7.3, Figure 7.4 and Figure 7.5).

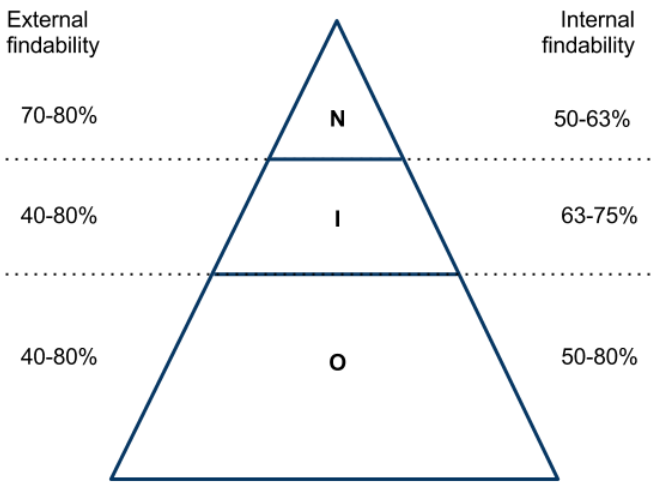


Figure 7.3. External and internal findability in ADL per level.

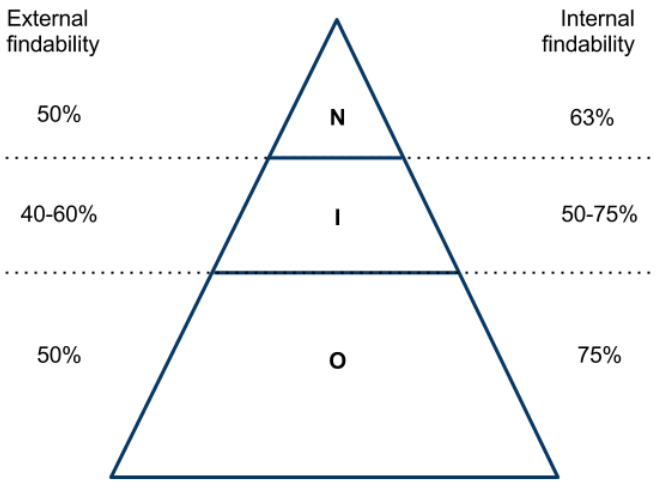


Figure 7.4. External and internal findability in KID per level.

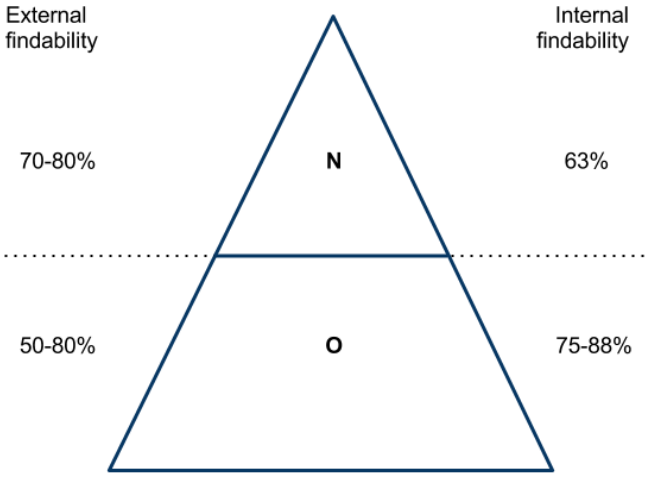


Figure 7.5. External and internal findability in Poma per level.

Generally both the external and the internal findability are good for the objects on the object level. There are no really low scores and the top values are in line with or better than the scores of the other levels. The specificity of the SAPs is not taken into account and the focus of the objects are more narrow on the object level than of the other two. The framework only takes the Access Target Area into account (see Chapter 3), not the meaning of the SAPs. In ADL and Poma the range between the lowest and highest external findability scores differ 30-40 percentage points and thereby the findability differs greatly between the individual objects. In KID the range between high and low scores are almost non-existent as no object has a PageRank value.

Because most of the objects are reachable internally by both link navigation and internal search, the scores for internal findability are better than the external findability scores, which largely

depends on the PageRank value of the object. As seen in Table 7.4, Table 7.5 and Table 7.6 the PageRank is the value that differs the most among the studied objects. The low top scores of the internal findability on the navigational level are due to the lack of indexing of the most general objects in the internal search engine. As the total findability scores are the sum of all six aspects the scores in many cases less extreme than the scores for external and internal findability. Only for objects, that has extreme scores, high or low, for both external and internal findability.

7.4 Findability and its impact on usage and navigation strategies

The navigation strategies use different aspects as shown in Figure 3.5 and Figure 4.3. For direct navigation reachability is the most important aspect. It must be possible to bookmark an object to be able to access it directly. Navigation by links also depends on reachability but web prestige is important too, as a high web prestige increases the probability that a user will encounter a link to the object. Reachability and web prestige, which forms the web presence, is also central for search engine navigation, but the most important aspect is the attributes of the object. The SAPs determine if the object is considered relevant by the web search engine. A high number of SAPs (a larger ATA) increases the probability of matching between the content of the SAPs and the search terms used by the searcher in the web search engine. A low external findability score means that the object's potential for all three strategies is worse than for an object with a high score.

It might be argued that a relationship between the level of internal findability and the length of the session could exist. A low internal findability score might indicate that the resource is hard to navigate in and that the users leave quickly. But short sessions might also indicate that users fullfill their needs without visiting a large number of objects. The navigation within a resource is harder to relate to the findability framework as the session indicators, e.g. length or levels visited, are more dependant on the user's need and motivation. The navigation to a resource is done in competition with other web resources, and the assessment of relevance is done before the user arrives at the resource and thus the aspects included in external findability might be more appropriate when discussing navigation and usage together with findability.

7.5 Reflections on the findability evaluation

The criteria in the evaluation have not been tested before. The evaluation is an exploratory attempt to evaluate or measure web findability in an ecological approach where user and system are integrated. The findability evaluation scheme should be tested and developed in future research. One way to study internal findability might be in experimental settings where users get pre-defined work tasks or search tasks and thereafter are free to navigate and search a specific site. Another is applying the present approach and criteria to different kinds of web sites. A third could be a test site on the web for A/B-testing. Different aspects could be altered over time and the

impact of the usage could be measured. But this would demand attractive content that would please real world users (Borlund, 2000). It would also be possible to look at the findability of objects in different parts of usage. For example to study a number of objects in different parts of the distribution of the number of times an object have been accessed, together with objects that have not been accessed at all.

The object attribute indicators (number of SAPs and full text) might be valued to low in both the external and internal findability scores. The number of SAPs should probably have a fourth category with 20+ SAPs (4 points) to differentiate the analysed objects in terms of amount of metadata. In the present version the full text only gets one point in addition to the SAP-points, maybe full text is worth two or even 1000 points. Indexing full text adds a number of additional ways of finding the object, e.g. phrase searching. The points could be given in a completely different manner. One example is given in Table 7.7, where 0-1-100-1000 is awarded instead of 0-1-2-3 for the graded aspects.

Table 7.7. Findability scores for Poma where the alternative points 0-1-100-1000 are given instead of 0-1-2-3 in Table 7.6. Fulltext is seen as a part of Object attributes and is not awarded separately.

Id	Object Attributes	Accessibility	Internal navigation	Internal search	Reachability	Web Prestige	External findability	Internal findability	Total findability	Alt. total findability
P-N1	100	100	1	0	1	100	301	201	302	502
P-N2	100	100	1	0	1	1000	1201	201	1202	1402
P-N3	100	100	1	0	1	100	301	301	302	602
P-N4	100	100	1	0	1	100	301	201	302	502
P-O1	1000	100	1	1	1	0	1101	1102	1103	2203
P-O2	1000	100	1	1	1	100	1201	1102	1203	2203
P-O3	1000	100	1	1	1	100	1201	1102	1203	2203
P-O4	1000	100	1	1	1	100	1201	1102	1203	2203
P-O5	100	100	1	1	1	0	201	202	203	403
P-O6	100	100	1	1	1	0	201	202	203	403
P-O7	100	100	1	1	1	100	301	202	303	503
P-O8	100	100	1	1	1	0	201	202	203	403
P-O9	1000	100	1	1	1	0	1101	1102	1103	2203
P-O10	1000	100	1	1	1	0	1101	1102	1103	2203

In Table 7.7, given the alternative points, the difference between high and medium findability is much clearer than Table 7.6. I do not regard any of the scores as low (a score below 100). The degree of findability is clearer for all three findability scores (external, internal and total). A fourth score is added, *alternative total findability*, which is the sum of the external findability score and the internal findability score. The alternative findability score stresses object attributes and accessibility more as they are counted twice. The only studied object that stands out in the two total scores is P-N2 which has high web prestige, and which places the object between those with medium and high scores. The alternative total findability score highlights that there are numerous ways of calculating the scores. In the next section different weightnings of the aspects are discussed, which opens up for other calculations.

7.6 Weighted findability aspects

Based on the usage of the different navigation strategies presented in Chapter 4 the aspects could be weighted differently than in the findability framework presented in Chapter 3. The most frequently used navigation strategy is search engine navigation, and regardless if the search engine is used for topical searching or navigation to a known resource, the number of SAPs are important. To strengthen the importance of the points awarded to the aspect "object attributes" this aspect could be weighted as two or three times as important as in the original framework. The re-weighting will for example highlight the "weaker" objects or types of objects with fewer SAPs in a more clear way so the resource manager, know where to focus the improvements in relation to visitors arriving by a search engine.

In

Table 7.8 the original normalised scores for ADL are compared to two new weights, object attributes x3 and web prestige x3. Web prestige is the most important aspect for link navigation. And increasing number of prominent links on the web to objects enable more traffic by links.

Table 7.8. Percentages of total external findability normalised scores depending on different weights in ADL.

Id	Original framework	Object attributes weighted x3	Web prestige weighted x3
A-N1	70%	56%	81%
A-N2	80%	75%	88%
A-N3	80%	75%	88%
A-N4	80%	75%	88%
A-I1	80%	75%	88%
A-I2	70%	69%	69%
A-I3	50%	56%	31%
A-I4	70%	69%	69%
A-I5	50%	56%	31%
A-I6	40%	38%	25%
A-O1	60%	50%	63%
A-O2	40%	38%	25%
A-O3	80%	88%	75%
A-O4	60%	75%	38%

When object attributes is weighted higher the lack of text on a page becomes clearer. The top page in ADL (ADL-N1) gets a lower rating because of the minimalistic design. First pages based on Flash-intro would get an even lower score. On the other hand the first page of the text version of The little mermaid (ADL-O3) obtains an increased score due to both a great number of SAPs and full text.

In the third column the web prestige is tripled. Here a low number of inlinks or missing PageRank from Google drags the score down. The second page of The little mermaid (ADL-O2 and ADL-

O4) has low web prestige due to their nature of a continuation of the first page (which people will link to as it is the start of the tale). More problematic is the page which links to the different versions of The little mermaid (ADL-I3), as it is neither strong in terms of object attributes (text) nor web prestige (links). As site owner it might be a good idea to add some additional text to the page in order to strengthen the first aspect, and then it will, hopefully, get more inlinks in the long run. In this case PageRank is used as an indicator of web prestige and the PR-value is inherited by pages that are outlinked, so the result of weak pages in the middle of a resource is that they might lower the findability of the objects beneath them in the resource.

The weighting of different aspects can highlight differences and weaknesses in both the structure and contents of the resource. In this sense the findability framework is flexible and can be used by site owners to analyse their resources and improve their weakest parts. The concrete weighting of aspects depends on the purpose of the findability analysis, and can be adopted according to different goals. It is also possible to study the impact of improvements if the findability analysis is done both before and after the changes. It should also be possible to study changes over time if deployed e.g. once a year.

7.7 Automation of the findability analysis

Findability measures forms alternative performance indicators e.g. relation to IR evaluation on the web. The six aspects in the findability analysis can be automated. All seven indicators evaluated in the present study can be measured in a similar manner to the present, manual evaluation. The number of SAPs on a page can be counted by indexing the page with a locally implemented web crawler, and then excluding the stop words before counting the number of content bearing words on the page. Whether a text is full text might be harder to identify. Often there is information in the link-text (anchor text) and in the context of the links to the objects, e.g. ”article in full text” or “facsimile”, but in the later example the full text might be available only in the form of pictures and is thereby not searchable as full text.

The compliance to the WCAG 2.0 accessibility guidelines should be possible to test against a WCAG-testing site through an API. Otherwise it ought to be possible to implement the most important aspects of the WCAG 2.0 guidelines in dedicated findability analysis software.

Identifying link paths should be possible through analysing the site map or by indexing all the links studied. If the objects are reachable through the internal search automatically submitting queries to the search engine is feasable, for example the title of each object (derived from the indexing of the objects for the SAP-analysis) and by comparing the contents in the SERP with the URLs from the indexed links.

Determining the uniqueness and stability of the URLs of the objects ought to be easy by formulating some formal rules for the analysis and then check the indexed URLs. By an API it

should be possible to extract PageRank values directly from Google, at least in a slow pace (due possible Google restrictions), for all indexed URLs.

As discussed automatic calculations of findability aspects are most likely possible. If the findability could be done automatically for a whole resource, then the scores for each level could be calculated in several ways, for example the mean together with the range, and the median for both external and internal findability.

7.8 Chapter summary

The first research question (RQ1) and the last two sub-research questions on findability were addressed in the present chapter: First: *How findable are resource and objects from the web (RQ1b)?* The objects in the studied resources are findable on the web. There are some findability issues that could be improved, to increase the external findability. In KID the secure protocol HTTPS is used instead of HTTP and thus Google has not given the pages in the resource any PageRank values. The implication of no PageRank is that the objects will most often be ranked after similar objects with a PageRank value in the results pages. This might not interfere with the behaviour of using a search engine for navigation to a known site, but for topical searches it might lead a great deal of potential users to other resources than KID.

Generally the amount of Subject Access Points (SAPs) per object is not low but the objects would benefit from a larger number of SAPs. Additional SAPs could be both in the form of descriptions and keywords visible in the web browser, and as objects specific metadata in the header of the HTML pages. This recommendation is not based on the results in form of points in the findability evaluation; it is based on the contents of the actual objects studied and the low number of SAPs needed for the highest score in the findability analysis. Especially for topical searches in web search engines it is important to have a number of different SAPs, with content bearing keywords. A careful and sensitive analysis of the referring could give clues of what keywords the users use in their queries in the referring search engines for enhancing the objects. But it should be noted that only the users who actually have arrived at the resource are present in the log files, all potential users who found other objects on the web with other queries are not presented in the logs.

The second sub-research question is: *How findable are objects within the resource (RQ1c)?* Generally the objects are findable within the studied resources. All studied objects were findable through the link structure. Although there were long link paths from the front page in the top to the cultural heritage objects in both ADL and KID, there was an increase of information scent all the way along the link path, so the users ought to have no problems following it. The internal search engines in all three resources are optimized for finding the cultural heritage objects and the informational objects, so it is often impossible to find the general objects on the navigational level through searching. This focus on the digitalized objects is in line with the goals of the

resources, i.e. to mediate the cultural heritage. But results in that the "normal" search behaviour of using site search in Google like manner is not possible.

RQ1 is *How findable are the heritage resource and their objects?* Based on the two sub-research questions the answer to the research question is that both the CH resources and their objects generally are findable. The degree of findability differs between the studied objects, but the overall level of findability was rather good. The studied objects were chosen to represent typical objects and to illustrate the framework for findability evaluation. Whether they are representative in a statistical manner is not known.

Three other alternatives of the findability framework have also been explored. One is awarding points to the graded findability measures in a completely other way, i.e. 0-1-100-1000 instead of 0-3 (Section 7.5). The second possibility is the change of weighting between the different aspects within the framework to highlight weaknesses in the resource in relation to different navigation strategies (Section 7.6). And the potential of automating the whole process of findability evaluation for studying findability in a large scale was discussed, for example evaluating every object within a resource, not just a sample of typical objects is another possibility (Section 7.7).

In relation to the ELIS framework the degree of findability of a resource might influence the probability that a user finds and incorporates the resource in her information source horizon. When incorporated in the information source horizon the resource might be a part a information pathway. Resources not included in the information source horizon of a user might still be a part of a information pathway as users searches different search services, e.g. topical searches in a search engine, and thereby finds the resources. Another possibility is that the user visits web sites within her information source horizon which links to the on beforehand unknown resources. For both alternatives the degree of findability is important, as it affects the likelihood of a visit to a resource.

8 The use of the cultural heritage resources

This empirical chapter focuses on the second research question*: How do users find and use the cultural heritage resources?*, and precisely on the four first sub-research questions:

- Which navigation strategies are used by the users to access the resources (RQ2a)?
- On what level in the resources do the users arrive (RQ2b)?
- How do they navigate within the resources (RQ2c)?
- How many objects do the users access in a session (RQ2d)?

In the log files several aspects are studied. First the navigation strategy used to arrive at the site, with an in-depth analysis of the queries used in referring web search engines, and an analysis of the referring web sites in general and Wikipedia in particular. Then the session length, arrival level, and session paths are studied, with a focus on the impact of the different navigation strategies. Is for instance the users' behaviour different if they arrive by a link compared to a search in a web search engine? The bounce rate is studied and the users' county of origin is also studied using the log files.

The aspects studied in the log files are used to describe of the usage and of the users. First the methodology of the log analysis is addressed, then the empirical findings from the logs of the three cultural heritage resources. In the end of the chapter conclusions are made and discussed on the sub-research questions.

8.1 How users access the heritage resources[27]

8.1.1 Navigation strategies

In ADL 44,352 sessions have been identified, in KID 22,667 and in Poma 30,557 sessions. In more than 50% of the sessions a search engine was used to navigate to the resource (Figure 8.1). In KID and Poma link navigation played an important part with around 30% of the referring traffic. Direct navigation is the strategy with the lowest frequency, used in 6-17% of the total number of sessions. The exact numbers are given in Appendix 15.

[27] In Fransson (2012) some of the results concerning KID were presented. Here the analysis has been developed and some of the results has been updated, for example the bounce rate.

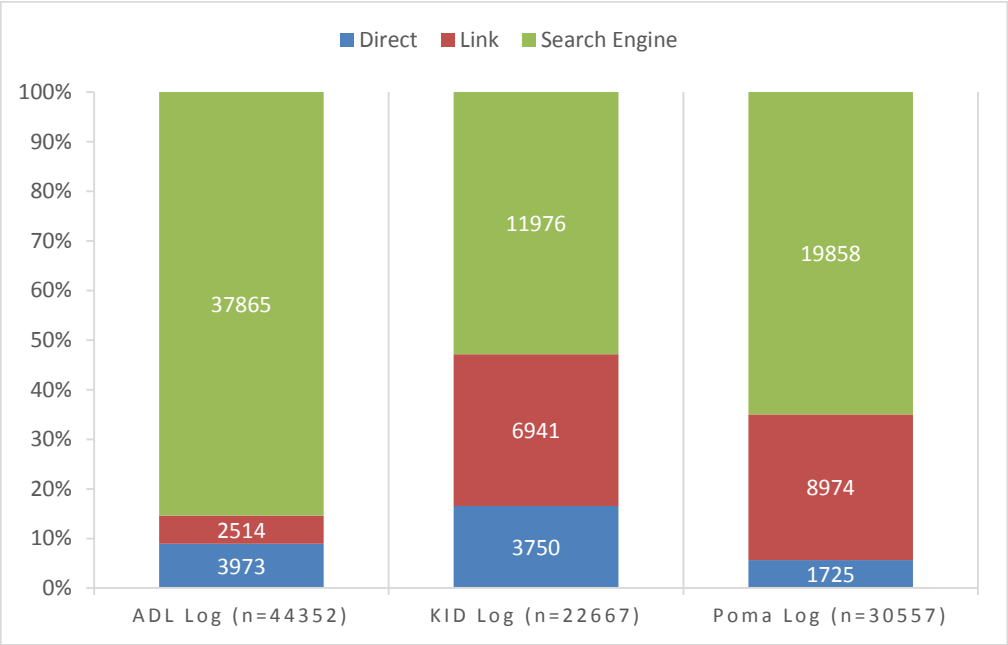


Figure 8.1. The distribution of the navigation strategies in the three resources.

The navigation strategies are further explored below with an indepth analysis of the referring search engines and web sites.

8.1.2 Distribution of referring search engines

Google's web search engine is the dominant search engine among the users of the three resources. As shown in Table 8.1 and Appendix 9 the reliance on Google web search engine is between 95% and 98%, plus some referring traffic from Google services like Images, Translate and Web cache. The second largest web search engine is Bing, with a share between 0.7% and 1.4%. Yahoo and its national versions of their search engine, including Altavista, is the third search engine with around 0.5-1% of the referring traffic. The Danish find.tdc.dk has a small share around 0.2% in the two Danish resources, and the search service uses Google's index.

Table 8.1. Distribution of the two largest referring web search engines (based on Appendix 9).

Web search engine	ADL	KID	Poma
Google	98,4%	96,5%	95,1%
Bing	0,7%	0,7%	1,4%
Other	0,9%	2,8%	3,5%

8.1.3 Queries used in the referring search engines

The samples of queries from the referring search engines were classified according to Table 4.1. In ADL 191 queries was analysed out of 19046 queries in the logs. In KID 57 queries was analysed out of 5652 and 49 queries out of 4846 in Poma. The queries are distinguish from the

number of times they occur. The distributions in Table 8.2 are based on the number of occurrences of each query type, i.e. the frequent top queries are counted a large number of times, and that is why the number of queries analysed from each log do not match the total number in the table. In Appendix 16 the queries in the samples from the whole distributions are listed as well as the 20 most frequent queries in the resources are listed, i.e. the head of the long tail.

In Poma the most frequent query "guaman poma" (670 occurrences of 741) is classified as navigational, but the query is ambiguous and could be seen as informational. The top query in KID is "weilbachs kunstnerleksikon" (571 occurrences of 642) which is interpreted as a navigational query to get to the resource, but a share of the searches might be about the resource in an informational meaning. In ADL the top query was "adl", which is a fairly typical navigational query. The distribution in Table 8.2 would have been completely different if the top queries had been classified as informational, or as belonging to both the categories. But based on their high frequency of occurrences compared to other queries it is logical to assume that the majority of the occurrences are navigational queries (see the Tables in Appendix 16).

In KID and Poma about 90% of the queries were navigational because of the top queries discussed above. In ADL the distribution was equal between the two types. There were no transactional queries in any of the samples, which might depend on how the category is defined. In Jansen et al. (2008) queries with "obtaining" terms, which are used to get hold of some specific information and includes terms like *lyrics* and terms for material type, e.g. *drawing*, were defined as transactional. But here the same terms were seen as specifications of informational queries.

Table 8.2. The distribution of informational, navigational and transactional queries in the query-sample (based on occurences).

	ADL #	ADL %	KID #	KID %	Poma #	Poma %
Informational	293	49%	70	11%	69	9%
Navigational	305	51%	572	89%	673	91%
Transactional	0	0%	0	0%	0	0%
Total	595	100%	642	100%	741	100%

Broders' original study the sample of 400 queries in an Altavista log contained 48% informational, 20% navigational and 30% transactional queries, based on the assumption that if the queries were not navigational or transactional they were informational (Broder, 2002). Rose and Levinson found a larger share of informational queries (~60%) in a similar study (Rose & Levinson, 2004). A third study found informational queries in 80% of the cases by using an automatic method (Jansen et al., 2008). In the studies the categories where defined slightly different definitions of the query intentions. Lewandowski found different distribution within different topic areas when studying query types in web search engine logs. For example where the distribution of informational queries around 50% in the topic area "Education or humanities", 70% in "Society, culture, ethnicity or religion", and under 10% in "People, places or things". Navigational queries were inversely proportional, but in a couple of topical areas transactional

had a share of around 30%, notably "Computers or Internet" and "Sex and pornography" (Lewandowski, 2006).

The queries categorised as informational were classified into subcategories according to the elements in the query. For example the ADL-query "herman bang chopin" is classified as *Creator* (Herman Bang is an author) and *Specific Object* (Chopin is a text written by Herman Bang). The result is shown in Table 8.3.

Table 8.3. The distribution of the informational subcategories in the query-sample (ADL n=293, KID n=70, and Poma n=60). The number of actual occurances of each subcategory is shown, not their share of the informational queries (see Table 8.4).

Informational subcategory	ADL #	ADL %	KID #	KID %	Poma #	Poma %
Creator - CR	163	56%	57	81%	0	0%
Specific Object - SO	120	41%	12	17%	16	23%
Full text - FU	27	9%	0	0%	5	7%
Geographic - GEO	4	1%	6	9%	13	19%
Genre - GEN	54	18%	13	19%	0	0%
Time Period - TP	23	8%	9	13%	0	0%
Specific Institution - SI	0	0%	2	3%	0	0%
Topic - TO	48	16%	7	10%	49	71%
Type - TY	9	3%	4	6%	10	14%
Other - OT	19	6%	9	13%	9	13%
Total number of informational queries	*467*		*119*		*102*	
Average number of different subcategories per query	*1.6*		*1.7*		*1.5*	

The most frequently used informational subcategory in KID is *Creator*, which is present in 81% of the informational queries. The other subcategories are much less frequently used, in up to 19% of the queries. *Creator* is the frequently used subcategory in ADl as well, in 56% of the queries, but the subcategory *Specific Object* is also frequently used in the queries (41%). In Poma the use of subcategories is different. *Topic* is the most frequently used subcategory, present in 71% of the queries. The different patterns in the use of informational subcategories reflects the type of content in the CH resources. The texts in ADL and the artworks in KID are often found through their creator as they often are better known than their works of art and literature. In Poma on the other hand, users not familiar with the Inca chronicle probably search on a topic covered in Poma and finds the resource.

The average number of subcategories per query is similar in all three resources (~1.6). The measure does not take the number of subcategories per query into account, just the number of different subcategories.

The use of polyrepresentation (Larsen et al., 2006) in the informational subcategory (using two or more of the subcategories in the query) is shown in Table 8.4 (see Section 4.1.2). The use of

several subcategories in a query might indicate a certain domain knowledge, at least enough knowledge to use different topical aspects in the query.

Table 8.4. Polyrepresentive informational queries (including two or more of the informational subcategories in Table 8.3).

	# Polyrepresentative informational queries	% of total number of informational queries
ADL	148	32 %
KID	39	33 %
Poma	27	26 %

All queries and the classification into subcategories are given in Appendix 16. *Creator* often occurred together with *Specific Object* in both ADL and KID, for example in the form of the authors' last name together with the title (or parts of it). In Poma the subcategory *Creator* was not used for "Guaman Poma" because the name of the creator is the same as the name of the resource and the query was classified as navigational.

Geographic terms, like Portugal or Denmark, and terms for *Time Periods* often specified topical searches. In the same way *Genre* terms often specified searches for a *Creator* and/or a *Specific Object*, for example an analysis (genre) of a specific text. The *full text* features in ADL and Poma was used to some extent (~8%), and then almost every time without other types of search terms in the query. There were to versions of the full text query, one used a longer phrase and the other one a couple of very specific words in combination.

The share of queries using polyrepresentation is close to 30 % in all three resources. It is interesting that the distribution of polyrepresentation in informational queries are similar in the different topics covered in the resources.

8.1.4 Referring links grouped per site

In the top referring sites, the referring links grouped per domain, and are listed in Appendix 17. Self-referring pages has been removed from the lists.

The top referrer in ADL is the site of the H.C. Andersen Center at the University of Southern Denmark (www.andersen.sdu.dk) which links to several Andersen pages in ADL. The top referrer in KID on the other hand is a privately owned portal (www.kunstonline.dk) about art in Denmark with links to artists, institutions, exhibitions, etc. KID is frequently linked to as a source of information in the artist biography section.

It is only among the referring sites of Poma that some social media sites are evident. The top three referring sites are blogs at blogspot.com and Facebook is on place number eight in the list with 95 referrals during the studied period. The top blog is referring 27% of all traffic to Poma, which is extreme and can only be compared to Google (38%).

Wikipedia is an important referrer to all three resources with three (KID and Poma) or four (ADL) versions among the 20 most common referrers. Danish Wikipedia is the most important language version in ADL and KID, and the Spanish in Poma.

In all three top 20 lists a site called search.conduit.com occurs. It is a site which uses Google index to deliver search results. For the user of the site it is like a regular web search engine, but to the analysis software it is a normal web site. This highlights the problem of identifying search engines automatically, and the referring links might equal, lesser used similar sites.

The differences in linking to the CH resources might explain the bounce rates connected to the link navigation strategy. Especially the high bounce rates in Poma might be explained by the traffic from the dominant blogs (see Section 8.2.4).

8.1.5 Referring Wikipedia pages

The referring Wikipedia pages are of different kinds of topics and from different language versions (see Table 8.5 for top 3 and Appendix 18 for top 10). In ADL the top Wikipedia pages are: *George Brandes* (author) from the Danish Wikipedia, *Thumbelina* (story) from English Wikipedia, and *Ludvig Holberg* from Wikipedia in Norwegian. On the top ten referring pages there are four different language versions of Wikipedia: Danish, English, Norwegian, and Japanese.

In KID the top referring Wikipedia pages are about the *Weilbachs Kunstnerleksikon* (Artist Encyclopaedia) in Danish, and two English pages on two artists. On the top ten referring pages there are three different language versions of Wikipedia: Danish, English, and German.

In Poma the top referring pages are the page about the chronicle in Spanish Wikipedia, and the entry about *Guaman Poma* in both the Spanish and the English versions of Wikipedia. On the top ten referring pages there are five different language versions of Wikipedia: Spanish, English, French, German, and Russian.

Table 8.5. Top three referring Wikipedia page in the studied resources (top 10 in Appendix 18).

	Rank	Wikipedia URL	#
ADL	1	http://da.wikipedia.org/wiki/Georg_Brandes	59
	2	http://en.wikipedia.org/wiki/Thumbelina	55
	3	http://no.wikipedia.org/wiki/Ludvig_Holberg	55
KID	1	http://da.wikipedia.org/wiki/Weilbachs_Kunstnerleksikon	115
	2	http://en.wikipedia.org/wiki/Vilhelm_Hammershøi	44
	3	http://en.wikipedia.org/wiki/Wilhelm_Freddie	37
Poma	1	http://es.wikipedia.org/wiki/Primer_Nueva_coronica_y_buen_gobierno	165
	2	http://es.wikipedia.org/wiki/Felipe_Guamán_Poma_de_Ayala	164
	3	http://en.wikipedia.org/wiki/Felipe_Guaman_Poma_de_Ayala	134

8.1.6 Users' countries of origin

For comparison between the resources the distribution of the session country of origin in the logs are shown in Figure 8.2, Figure 8.3 and Figure 8.4, with more than 1% of the traffic according to the analysis in Web Log Storming (Appendix 19).

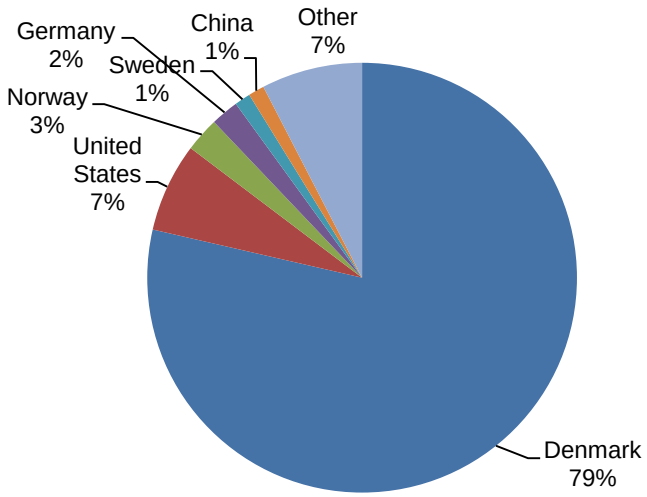


Figure 8.2. ADL users' country of origin in the logs.

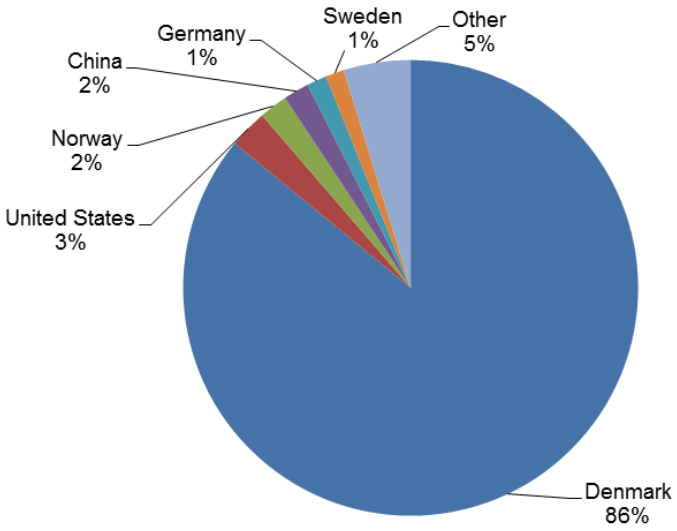


Figure 8.3. KID users' country of origin in the logs.

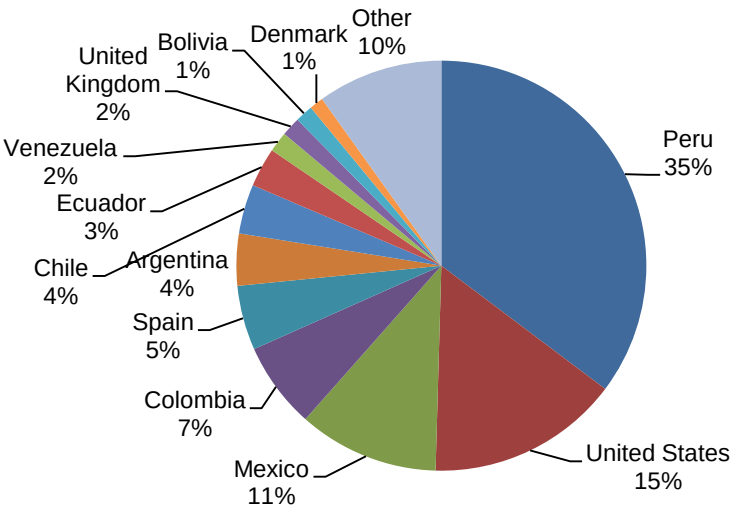


Figure 8.4. Poma users' country of origin in the logs.

In both ADL and KID most sessions has their origin in Denmark, as they are tightly linked to the Danish language (ADL) and Danish institutions and their holding of the physical objects (KID). In Poma 70% is from Spanish speaking countries, even more if counting the countries hiding in "other", plus the Spanish speaking population in the United States. This reflects the fact that Poma is an important historical document in Spanish.

8.2 The sessions

8.2.1 Where do the users arrive?

Two different session measurements were derived from all the log data covering three month analysis window: session length measured in number of pages viewed and the URL of the arrival page. The URLs of the arrival pages were divided into three groups according to the site structure. The average session length was compared to both arrival level and navigation strategy (Table 8.6).

Table 8.6. Number of sessions distributed on Navigation strategy and Arrival level.

		Direct #	Direct %	Link #	Link %	Search Engine #	Search Engine %	All strategies #	All strategies %
ADL	N	2145	5%	791	2%	1851	4%	4787	11%
	I	643	1%	1121	3%	28526	64%	30290	68%
	O	1185	3%	602	1%	7488	17%	9275	21%
	Total	3973	9%	2514	6%	37865	85%	44352	100%
KID	N	3125	14%	1878	8%	2486	11%	7489	33%
	I	458	2%	3649	16%	8523	38%	12630	56%
	O	142	1%	1447	6%	958	4%	2547	11%
	Total	3725	16%	6974	31%	11967	53%	22666	100%
Poma	N	1105	4%	8413	28%	4212	14%	13730	45%
	I	-		-		-		-	
	O	620	2%	561	2%	15646	51%	16827	55%
	Total	1725	6%	8974	29%	19858	65%	30557	100%

The most common arrival level is I, the middle level with artist or author information, in ADL and KID. In Poma which lack the I level, the object level is the most common arrival level. Except when the users arrive by direct navigation, they arrive at the upper N-level due to the general nature of the navigation strategy. Surprisingly few arrived at the object level (O) when using a search engine for navigation. Three reasons may be: the lack of metadata connected to the digitalized objects, the objects might be in fulltext and thereby highly specific texts, or how the users formulate their queries (see Section 8.1.3).

8.2.2 How long do the users stay?

The number of objects viewed is used for measuring the length of the visits, that is objects on all three levels not just on the object level (O) (the results are shown in Table 8.7).

Table 8.7. Average number of page views per session based on Navigation strategy and Arrival level.

		Direct	Link	Search Engine	All strategies
ADL	N	24.2	14.5	8.9	16.7
	I	16.5	12.2	3.2	3.8
	O	21.7	14.9	3.6	6.7
	Total	22.2	13.6	3.5	5.8
KID	N	15.3	15.4	15.7	15.5
	I	7.3	6.0	4.6	5.1
	O	6.8	12.2	4.7	9.1
	Total	14.0	9.8	6.9	9.0
Poma	N	17.4	2.4	8.4	5.8
	I	-	-	-	-
	O	16.5	3.2	1.8	8.3
	Total	17.1	2.5	3.2	7.2

How long the users stay on the site depends on the both which navigation strategy they use and on which level they arrive on. Direct-sessions are the longest in average. The sessions based on search engine navigation the shortest, except in Poma where the link navigation sessions are on average the shortest due to a high bounce rate (see Section 8.2.4). The difference in average session length is greater if analysed from the arrival level. The N-sessions are more than double as long as the I-sessions, on average, in both ADL and KID. The I-sessions are short. The I-sessions are users arriving at the resource at an informational object, at a topical relevant page, and one might assume that the links from the page leads to other topical relevant pages, and the user may therefore explore the resource further. The O-sessions are longer than the I-sessions, between 6.7 and 9.1 pages long on average in the three resources. The object the user arrives at the O-level are more narrow in topic than the objects at the I-level, and the surrounding objects at the O-level might be of a different topic.

Figure 8.1 illustrates a typical distribution of the session lengths.

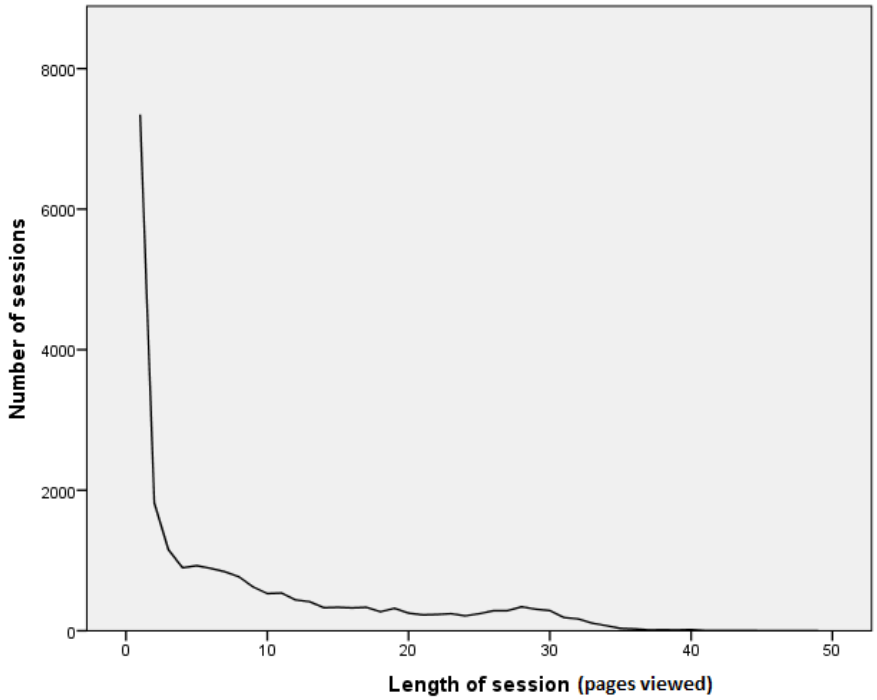


Figure 8.5. The frequency of session length in KID (n=22666) – a typical distribution.

The average length of the sessions indicates a group of users which looks at a large number of objects (on all levels). The users penetration of the resources is covered in the next section.

8.2.3 Where do the users go within the resource?

The length of a visit in a resource only measures how many objects a user looks at, not the type of objects looked at. To examine which levels within the resource the users visits, as discussed in Section 4.2.4, there are different path types within a resource based on the resource model. The distributions of the path types are listed in Appendix 20 and illustrated in Figure 8.6.

The session path types have different characteristics. In path types N1, I1 and O1 only one page is viewed, the arrival page (Figure 8.6). In all other types of paths at least two pages are viewed. The arrows in the figure only illustrate the levels visited in each path type, e.g. in N2 any number of pages on levels N and O might be viewed, and in any order besides the arrival at the N-level. In nine of the 15 session paths the user views at least one page on the Object level, which means they look at digitalized cultural objects. The paths with Object viewing are: N3, N4, I3, I4 and O1-O5.

The paths in ADL is described in Figure 8.6. In only 8% of all the sessions all three levels within ADL was visited (path types N4, I4 and O4). In these sessions most objects were visited, on average 26.5 objects per session.

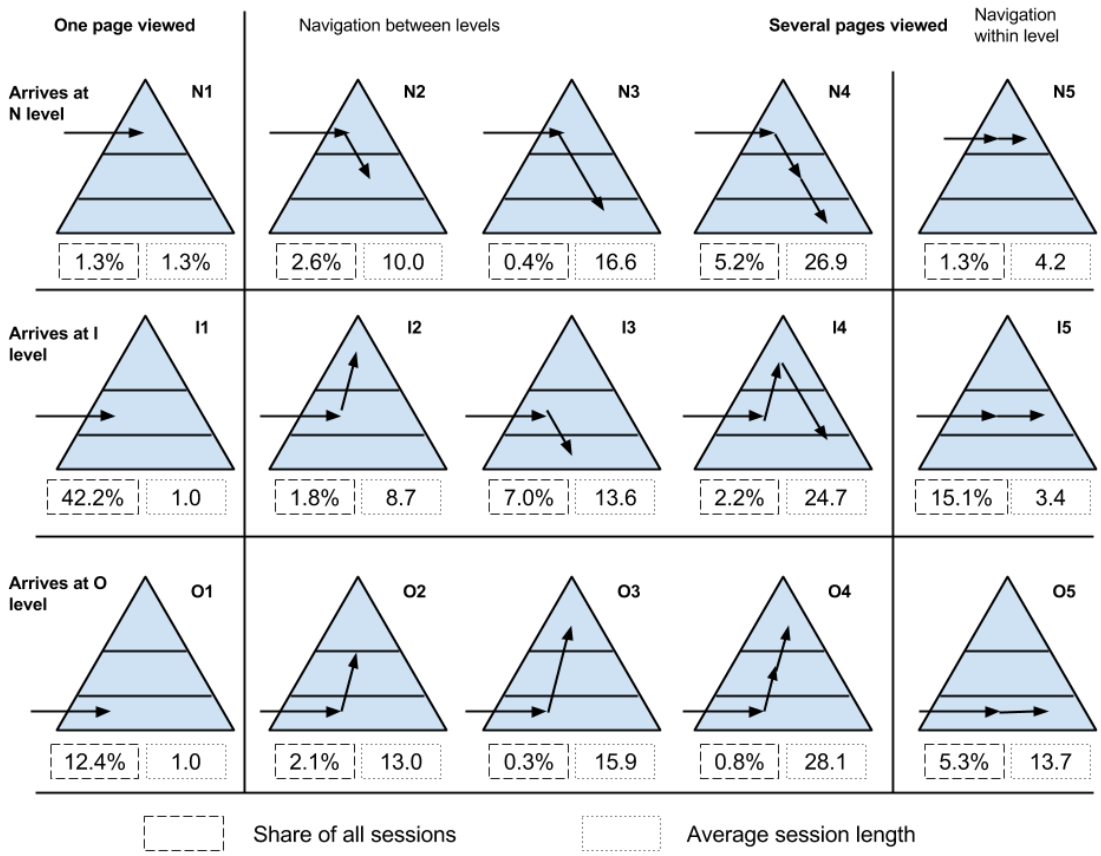


Figure 8.6. The 15 types of paths based arrival level, visited levels and number of pages viewed with the share of each path in ADL. The arrows are only illustrations of the arrival level and the levels visited thereafter in the session. The two values under each path are the share of all sessions (left) and the average session length (right).

Object pages are viewed in 36% of the sessions (in 9 of the 15 paths in Figure 8.6, path types N3, N4, I3, I4, O1-O5) in ADL. In the sessions which include viewing at least one O-level page the average session length is 12.2 viewed pages. The average number of pages viewed in the sessions where no O-level pages are viewed is just 6.5 page views.

The paths may be seen as parts of the ELIS information pathways discussed in Section 2.2. The information pathways consist of paths within and between different resources. Some of the paths in Figure 8.6 can be interpreted as berrypicking behaviour (Section 2.5). Especially the path where the user narrows her focus and looks at more specific objects as the search process progress, i.e. N3, N4, I3, I4 and O5.

The results for KID is shown in Figure 8.7. In KID only 28% of all the sessions all three levels within KID was visited (path types N4, I4 and O4). In these sessions most objects were visited, on average 18.6 objects per session.

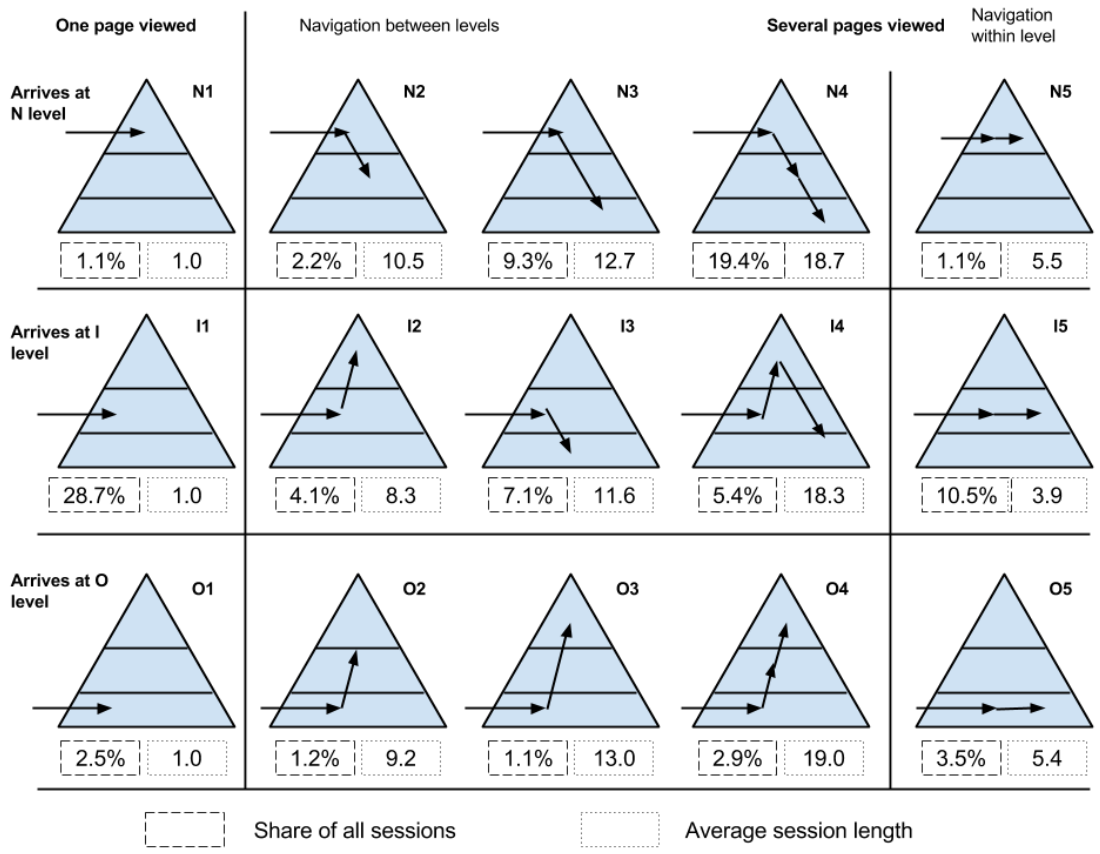


Figure 8.7. The 15 types of paths based arrival level, visited levels and number of pages viewed with the share of each path in KID. The arrows are just illustrations of the arrival level and the levels visited thereafter in the session. The two values under each path are the share of all sessions (left) and the average session length (right).

Object pages are viewed in 52% of the sessions (in 9 of the 15 paths in Figure 8.7, path types N3, N4, I3, I4, O1-O5). In the sessions which includes viewing at least one O-level page the average session length is 14.8 viewed pages. The average number of pages viewed in the sessions where no O-level pages are viewed is just 2.8 page views.

Poma is only divided into two levels and therefore the 15 possible types of paths in Figure 8.6 and Figure 8.7 are reduced to six in Figure 8.8. In only 17% of all the sessions both levels within Poma were visited (path types N4 and O4). In these sessions most objects were visited, on average 15.1 objects per session.

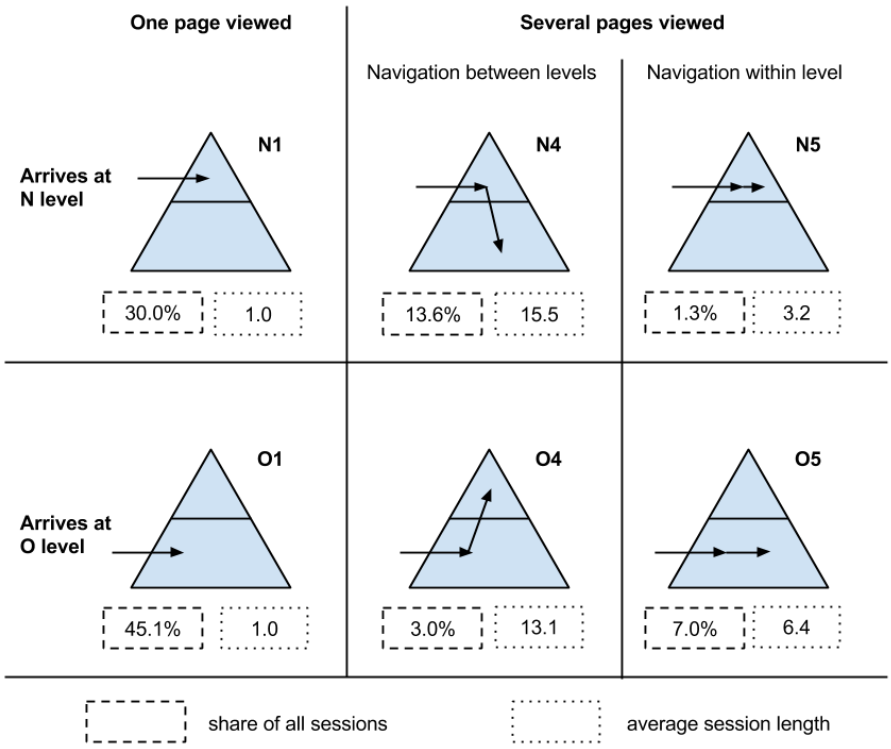


Figure 8.8. The 6 types paths in a two Poma-level version of Figure 7.4. The paths are based on arrival level, visited levels and number of pages viewed. The arrows are just illustrations of the arrival level and the levels visited thereafter in the session. The two values under each path are the share of all sessions (left) and the average session length (right).

Cultural heritage objects, in this case digitalized pages from the chronicle, was viewed in 69% of the sessions (path types N4, O1, O4 and O5), but only 25% of the sessions were longer than one page view. The average length of the sessions where at least one cultural heritage objects viewed is 4.9 viewed objects, to be compared with 1.1 objects viewed in the sessions were no cultural objects are looked at. The high percentage of CH objects viewed compared to ADL and KID might be explained by the lack of an informational level.

Almost 14% of the users arrive at the navigational level and continues down in the resource and also looks at least one digitalized page. But just 3% of the users arriving at an object (O-level) also visit the N-level. For users arriving at the N-level there is a probability of 33% that they look at more than one object. But for users arriving at the O-level the probability is just 18% that they go beyond the first page.

CH objects are viewed in 36% (ADL), 52% (KID) and 69% (Poma) of the sessions. The digitized CH is accessed, on average, in 50% of the sessions, which is good regarding all bouncing users in the navigational and informational levels. This should also be compared to the degree in which the N and I levels are visited. The navigational level is visited in between 17% (ADL) and 47-48% (KID and Poma), which is on average lower than the O-level. The informational level is visited in 64% of the sessions in ADL and in 74% of the KID sessions, which is significantly

higher than the average on the O-level. Based on these measures is the information about the CH objects on the I-level more used than the digitized CH objects.

The main limitation of the analysis of paths through the resource model is that it is not possible to discuss navigation or search strategies in a more complex manner than levels visited. Canter et al. (1985) uses six measurements of users' navigational behaviour in computer based systems to describe the users' behaviour in the terms of five search strategies: scanning; browsing; searching; exploring; and wandering. The strategy describes the navigation within the system with the first page as the starting point. A similar approach might be possible with the present dataset and it would be interesting to move closer to actual search strategies.

Another analysis that might further the understanding and the navigation within the session is a probability calculation of the transitions between the different levels. To investigate the percentage of the users visiting a page on the navigational level that goes to another N-level page and the percentage that goes on to an I-level page, etc. The calculation could result in Markov chains with four states (N, I, O and Exit) for both the level and the navigation strategy (Borges & Levene, 2007). Through, such an analysis other patterns might be discovered.

8.2.4 How many leaves directly?

The bounce rate is a measure of how many users that only visit one page before they leave the site. Bouncing is ambiguous because it describes a behavior which might be based on completely different judgments. The user could be completely satisfied with the information on the first page or the site did not correspond at all to the user's need.

The bounce rate on different levels and different navigation strategies might indicate different types of information needs. The phrases in the referring search engines shows that between 10% and 50% of the users arriving by a web search engine has an informational need (as expressed by the type of query in Table 8.2). Those users probably do not leave immediately, nor do the users arriving through direct navigation (a bounce rate between 2% and 7% in Table 8.8, Table 8.9 and Table 8.10).

Low bounce rate indicates a high correspondence between information need and the information at the level. The users continue their visits by looking at other pages (objects) to a large degree. But a low bounce rate could also mean that the site is hard to find and only the most interested and motivated users find their way to the site.

High bounce rate can also indicate that the landing page is perceived as non-relevant by the users and they therefore leave the site immediately, which is the common interpretation of bounce rate. This might happen if the user's query in the referring web search engine is ambiguous or unclear. Another possibility in search engines is that the representation in the results page (SERP) in not matching the actual content on the page or site. But high bounce rate can also indicate a fulfilment of the information need directly. If arriving at an object with the desired information when landing

at the site, visiting other objects might not be relevant for the user. In this sense high bounce rate is good.

It is interesting to measure of how users respond to different parts of a site. The three resources have overall very different bounce rates. The total bounce rate (all levels and all strategies) ranges from 32% in KID to 75% in Poma, with ADL placed in between at 56%. A general tendency is that the bounce rate is lowest at the navigational level (at the top of the resource). At the other levels the bounce rate is higher, in ADL it is around 60% at both I- and O-levels. In KID the bounce rate is 52 at I-level and only 23% at O-level.

Table 8.8. Bounce rate distributed on Navigation strategy and Arrival level in ADL (n=44352).

	Direct	Link	Search Engine	All strategies
N	0.4%	19%	22%	12%
I	5%	29%	64%	62%
O	3%	20%	72%	59%
All levels	2%	24%	64%	56%

Table 8.9. Bounce rate distributed on Navigation strategy and Arrival level in KID (n=22666).

	Direct	Link	Search Engine	All strategies
N	1%	6%	4%	3%
I	48%	43%	55%	52%
O	2%	19%	31%	23%
All levels	7%	28%	43%	32%

Table 8.10. Bounce rate distributed on Navigation strategy and Arrival level in Poma (n=30557).

	Direct	Link	Search Engine	All strategies
N	4%	90%	36%	67%
I	-	-	-	-
O	7%	71%	85%	82%
All levels	5%	89%	75%	75%

The navigation strategies also have an impact in the bounce rates. Users navigating through direct navigation bounce to a small degree (2%-7%). The rate for search engine navigation on the other hand is generally high (43%-75%), which is in line with previously observed pogo-sticking behaviour of search engine users (Thurow & Musica, 2009), where users jump between the search engine results page and the links in the list.

The bounce rate is moderate (24% and 28%) in ADL and KID when users navigates by links, but in Poma the rate is 89%. This high rate might be caused by the traffic from three blogs instead of more traditional cultural heritage sites as in the case of ADL and KID. Maybe the users bounce because of the change of genre, from the blogosphere to historic document in full text.

In KID the bounce rate is 32% in total, which might be considered low on the web. Both the navigation strategies and the arrival level display great differences. The Direct-navigators have a low bounce rate, mainly because they are returning to a known site. Search engine navigation has the highest bounce rate, just fewer than 43%. The bounce rate differs more between the arrival levels. The N-level has a rather low rate on 3% in average and there is no large difference between the different navigation strategies. On the other hand the I-level has a bounce rate of half of the visitors, which is a surprise. The reason why is unclear. Maybe the users were looking for something else. On the O-level there is a big difference between the bounce rates of the different navigation strategies. Direct navigation has a rate of 2% and search engine navigation 31%. For comparison the bounce rate was 65% in Europeana, the European cultural heritage search service (Clark, 2011).

8.3 Chapter summary

As stated in the beginning of the chapter the four sub-research questions have been addressed. The first sub-research question is *Which navigation strategies are used by the users to access the resources?* Of the three navigation strategies search engine navigation is used most frequently in the logs. By examining the queries in the referring web search engines by a sample of every hundredth query a description emerges of the types of information need the users' had. Navigational queries dominated the sample because the frequent use of the first queries in the sample, i.e. "guaman poma", which in all three resources were categorised as navigational queries. The distributions between the navigational and informational queries in the samples are probably not representative. More interesting is the structure of the informational queries where circa 1.6 element types were combined in each query. Often a name of the creator was combined with the parts of the work. Another pattern was that geographical or time terms were used to specify topical searches. The users who followed links arrived from different types of sites. Wikipedia is one large referrer, both the Danish version and other language versions. Social media sites are quite absent as referrers. Facebook is a top 20 referring site in two of the resources, and in Poma three blogs generates a lot of visits, but otherwise the users followslinks from more "traditional" sites on a similar topic.

The next question addressed was *On what level in the resources do the users arrive?* The users arrive at different levels in different resources. In ADL the large majority of users lands at the informational level containing information about the authors. The KID users also arrive at the artist information in most sessions, and very few arrive at some artwork at the object level. In Poma both the navigational level and the object level are frequent landing places, and due to the nature of the resource there is no informational level in the middle. Generally the direct navigator arrives at the navigational level, which is natural if you type in the URL or uses a saved bookmark (if not having a special interest in for example a specific author). Both the link navigators and the search engine navigators most frequently arrive at the informational level where there detailed

information about authors and artists (in ADL and KID) exits. In Poma they most often land at the object level, but there is no great difference between neither the strategies nor the arrival level. Based on the patterns described above it is clear that the information about the cultural heritage objects, about the authors, the artists and the museum, is the content that attracts the users, or at least leads them to the cultural heritage resources.

After the users have arrived, *How do they navigate within the resources?*, that is, sub-research question 1b. The different session paths reveals several patterns. Cultural heritage objects at the object level is viewed in 36% of the sessions in ADL, in 52% of the sessions in KID and 69% in Poma. In approximatly half of the visits no cultural heritage objects are viewed, just information about them. On the other hand a large share of the session is just one page view. The users "bounce" in 75% of the sessions in Poma, 56% in ADL and 32% in KID. The bounce rate is lowest in the group who arrives at the navigational level. They most often look at more objects. This is one answer to the next sub-research question, *How many objects do the users access in a session?* The session length, the number of objects viewed during a session is used to study the length instead of the time. In KID the average number of objects viewed in a session is 9.0, in ADL the average number is 5.8 and only 3.7 in Poma. The average session length is inversely proportional to the bounce rate; a large number of short sessions lower the average length.

In a related study of the usage of Europeana based log file analysis the CIBER research group identified different user types, or at least usage patterns: *user*, *one shot*, *mobile*, and *heavy*. In the *user* category are the majority of users and they are the users not belonging to any of the other categories. *One shot* are the bouncers, *mobile* are users using a mobile device regardless of the length of the session, and in the *heavy* category users who are probably working with the development of Europeana (Clark, 2011). The four usage patterns used by the CIBER team are unambiguous. In the present research the users are not divided into categories. In the log analysis the navigation strategy used to get to the site is one way of looking at the users; direct navigators, link-followers, and search engine users. Another focus has been on the usage and on the session paths and the length of the sessions.

In relation to the ELIS framework it is not possible to determine if the visits are for hedonic or utilitarian purposes, if the context is everyday life, work or education. The referring URL reveals the previous site visited when the user has navigated through links or a search engine. When the referring URL was a site it is a part of the information pathway of the user, but when it is a search engine the pathway is obstructed as the search engine is used as a tool to get to the next resource. Direct navigation indicated that the resource is known by the user and a part of the closest zones in her information source horizon.

9 User characteristics

In this chapter two sub-research questions will be addressed: *Which demographics characterize the users (RQ2e)?* and *Why do users visit the resources (RQ2f)?* The questions are answered through the survey data, which is presented and analysed. First the characteristics of the respondents are presented, which is followed by an analysis of their use of navigation strategies. Two focuses are the respondents' intention with their visit and in which context the visit takes place. The analysis of the survey describe the usage and the users. In the end of the chapter conclusions are drawn and discussed in relation to the sub-research questions, as well as a comparison of the results of the survey and the logs files. The survey questions are presented in Appendix 7.

9.1 Characteristics of the Users

The characteristics of the participants in the survey are displayed in Table 9.1, and are presented for two reasons. First to give an overview of the demographic data, create a picture of the users, or at least of the participants in the survey. The second reason is that the main survey questions have been tested against the characteristics (except for country of origin). Any statistical significant relationship between demographic characteristics and factors like navigation strategy or task context is presented under each heading.

The respondents were slightly more males in all three resources (55%-60%). The ADL-respondents were on average older than in the other resources. Both KID and Poma had a larger group of younger (19-30 years) respondents, whereas ADL had a large group of older respondents (66+).

Table 9.1. The characteristics of the participants in the survey.

Demographics statistics		ADL n=55	ADL %	KID n=256	KID %	Poma n=44	Poma %
Gender	Female	22	40%	114	45%	20	46%
	Male	33	60%	142	56%	24	55%
Age	-18	0	0%	0	0%	2	5%
	19-30	17	31%	15	6%	11	25%
	31-45	13	24%	46	18%	16	37%
	46-65	16	29%	125	49%	13	30%
	66-	9	16%	70	27%	2	5%
Country of origin	Denmark	42	76%	229	90%	3	7%
	Other country	13	24%	27	11%	41	93%
Level of Web search skills	High	28	51%	134	52%	22	50%
	Low	27	49%	122	48%	22	50%

The level of web search skills is based on three questions in the survey where the participants rate their knowledge and skills around information seeking on the Internet. The distribution between perceived high and low web search skills is the same in all three resources, close to 50%-50%. In KID there was a tendency that the oldest respondents rated their web search skills as low to a greater extent than the other age groups. The tendency was weaker in the 46-65 years group, but still present. In KID there were also more men than women that rated their search skills as low. Whether or not the tendencies reflect lower skills or if it is about confidence concerning the medium is impossible to say. The low web search skills among elderly and men correlates with the fact that there appears to be a large user group of older male hobbyists (see Section 9.4). No clear tendencies were found in ADL or Poma, perhaps due to the small number of respondents.

9.2 Navigation strategies

The three navigation strategies discussed above in Section 4.2.3 on were split into five in the survey to capture how the search engine was used for navigation and the type of direct navigation. The search engine answers reflect two types of searches: informational and navigational (Broder, 2002). Figure 9.1 presents the result for the three CH resources (based on the data presented in detail in Appendix 22).

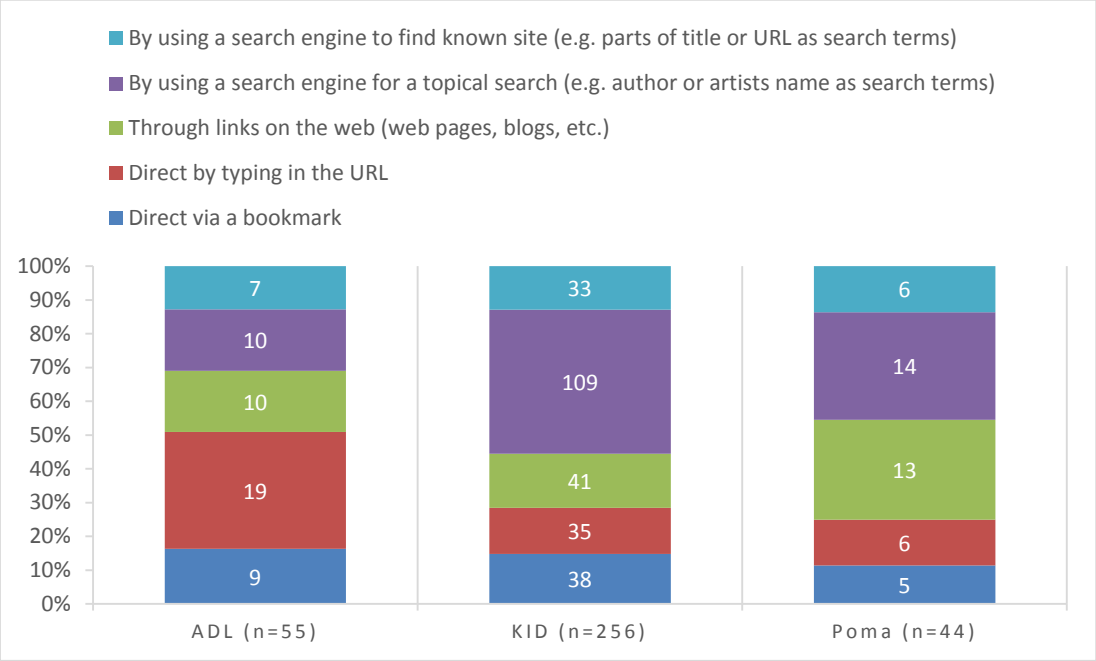


Figure 9.1. Staple diagram of answers to the survey question "How did you reach this site?".

The ADL respondents were typing in the URL more often than the other respondents (35%). Perhaps because of the easiness of typing just six characters: adl.dk. But typing in the URL is not at the expense of the other go-to-a-known-site strategies; direct via a bookmark or using a search engine for locating a known site. The usage of the navigational strategies is stable across the resources (11%-16%).

Topical searching in a search engine is widely used by the KID-respondents (43%), and link navigation by the Poma-respondents (30%). Combined the topical search strategies, link navigation and topical search in a search engine are more widely used in KID and Poma (around 60%) than in ADL (below 40%).

In KID there is a significant relationship between the use of navigation strategy and the level of web search skills (χ^2=10.092, df=4, $p<0.05$). The group with low skills navigated to the site through a bookmark and through links to a larger extent. The high web search skills group more often typed in the URL and arrived by a topic search in a search engine.

The use of navigation strategies might also be discussed in relation to the information source horizons of the users (Savolainen, 2008). The direct navigation strategies is probably more frequently used in the zones close to the user (zones 1 and 2 in Figure 2.4), when the resource is well known and regularly visited. Resources in a more peripheral zones (zone 3 and beyond in Figure 2.4) might not be regularly visited and the user has presumably not stored the URL as a bookmark or does not remember the URL. The use of navigation strategies might also depend on the user's stock of knowledge in relation to the search task at hand (Figure 2.3) or on intervening factors in the context (Figure 2.2).

9.3 Users intentions and task contexts

The task context measured in the survey is a combination of work task, information seeking task and information searching task together with the non-work task concept of coincidence. ADL and KID are mostly visited for hobby and leisure reasons (~46%), and KID secondly visited for work purposes (33%) and ADL for both work and school and study contexts (26%). Poma is primarily used in a school and study context (50%) and secondly for hobby and leisure (27%) according to the survey (Figure 9.2 and Appendix 22).

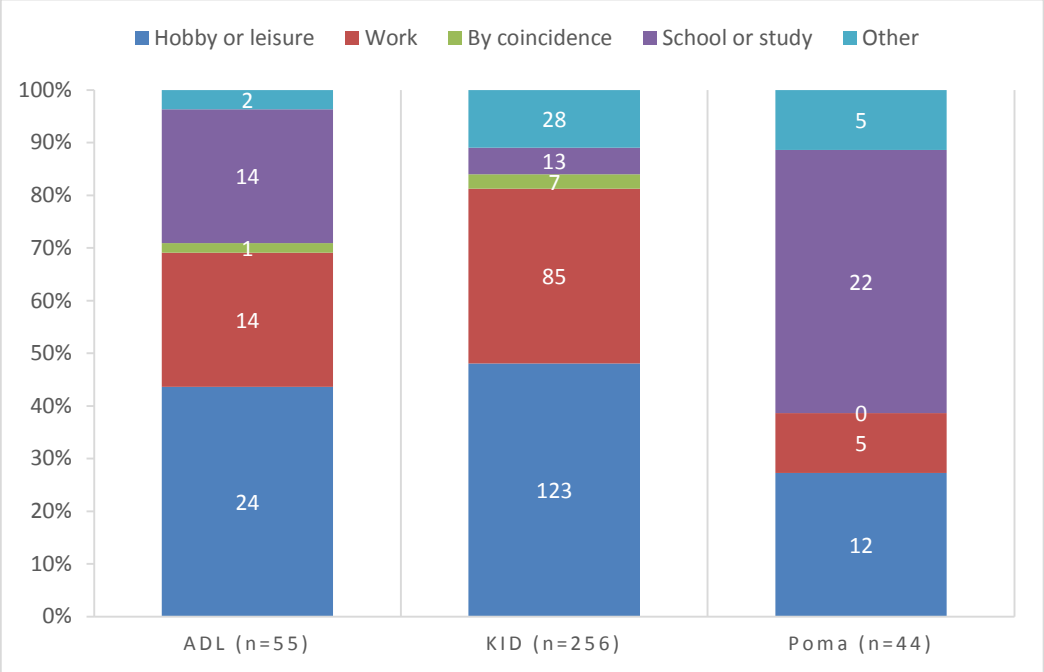


Figure 9.2. Staple diagram of answers to the survey question "In what context do you visit the web site?".

The contexts of the visits are a mix of hedonic and utilitarian. Almost half of the visits are for hobby or leisure reasons. The interpretation is that the CH resources are used in the everyday life of the users, besides for study or work purposes, which is a goal with the digitization. Due to the highly specific topic of Poma it is not surprising that it is mainly used in a study context, as it is a key text in university courses in both North and South America.

The question about the intention with the visit when respondents answered the survey could be answered with multiple answers, in contrast to all other questions where just one answer could be chosen. On average there were 1.6 answers per respondent in ADL, 1.8 answers in KID, and 1.2 answers on average by the Poma-respondents (see Figure 9.3 which is based on the data in Appendix 22).

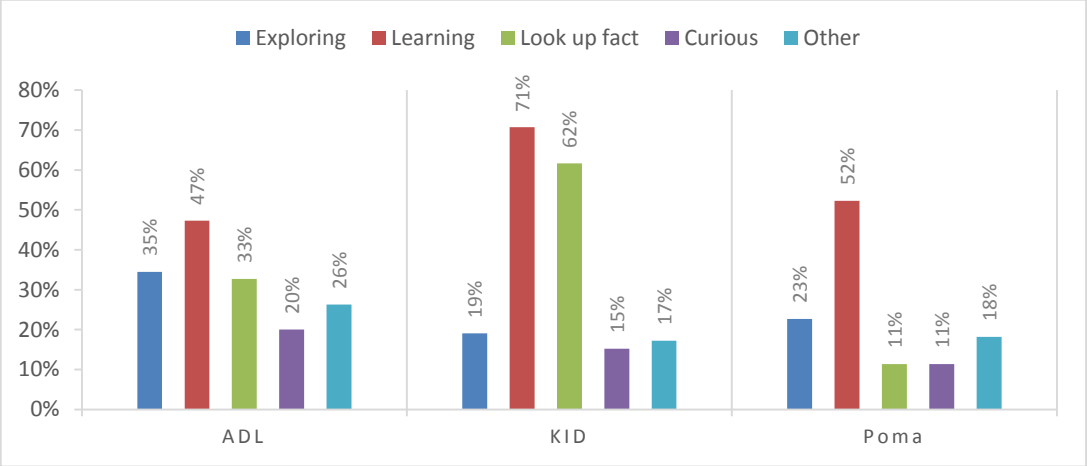


Figure 9.3. Staple diagram of answers to the survey question "Why are you visiting this resource today?".

Learning is the dominant reason for the visit in all three resources with an answer frequency between 47% and 71%. Learning is an important intention in several contexts, not only in the school or study context. In ADL *Exploring* and *Look up fact* was given as answer by a third of the respondents. In KID 62% was looking for facts as the second most frequent intention with the visit. *Exploring* was the second most frequent purpose in Poma (23%).

Exploring, *learning* and *curious* are explorative searches (Figure 4.2), in contrast to the intention of answer a question or look up a fact (White & Roth, 2009). The respondents answering *curious* might also be motivated by hedonic casual-leisure needs (Elsweiler et al., 2011). The diversity among the intentions of the visits indicates that the CH resources corresponds to both simple and complex information needs.

9.4 The users and their usage in the resources

The statistical testing on the survey answers is carried out to discover patterns within the dataset, not on the data as a sample of a random population. It is used as a tool to explore the respondents' answers. In KID there were a large number of respondents (n=256) and statistical test were done on the data set. On the ADL (n=55) and Poma (n=44) datasets of survey answers no statistical testing were done due to the small datasets. The test values below has no statistic relevance outside the tested dataset, and it is not used for statistically based generalisations.

In ADL a majority of the respondents are over 30 years old and 90% are from Denmark. According to the distribution of the navigation strategies a majority knew where they were going, that is to more than 60% of the respondents ADL was a known site and they navigated directly to it. The most frequent context of the visit was *hobby or leisure*, and the contexts of *work* or *school or study* shared the second place. The main intention with the visit was *learning*, but also *exploring* and *looking up fact*.

In Poma the respondents' primarily visited Poma in a *school or study* context, followed by the context of *hobby or leisure*. The most frequent intention with the visit was *learning*, and the second most common purpose was *exploring*. Many of the respondents navigated through *links* (in blogs) or by *topical searching in search engines*. The majority of the respondents were under 45 years old and 68% were living in South or North America (3 respondents lived in Denmark).

In KID there is a strong relationship between the task context and the navigation strategy (Appendix 23: χ^2=61.032, df=16, p<0.01). *Direct navigation by typing in the URL* is most frequent in a *work* context (80%). *Hobby or leisure* is the most common context for several other navigation strategies: *links* (61%, *topical search in search engine* (57%), and *search for a known site in a search engine* (55%).

There is also a relationship between task context and frequency of visit (χ^2=55.736, df=12, p<0.01). Users in a *work* context have often visited the site more than five times (80% of the users in a work context). Participants in a *school or study* context are to a large extent new visitors or have just visited the site a couple of times before. Most of the *coincidence*-visitors are visiting the site for the first time.

Work task was cross tabulated with gender and there is a significant relationship (χ^2=18.676, df=4, p<0.01). *Females* visit the site with a *work* or *study* purpose, while the *males* more often access the site in a *hobby or leisure* context. There is also a statistical significant relationship between age group and task context (χ^2=51.826, df=12, p<0.01). Within each age group the most common task context is clear. In the two younger groups the *work* context is dominant. And in the older groups the context of *hobby or leisure* is the most common (in Appendix 23).

The KID users were asked why they visited the resource and they were given the choices: *exploring*, *learning* and *look up fact*, based on the concept of exploratory search together with *curious* and *other*. In this particular case it was possible to choose more than one answer. The most common intention was *learning* (71%), but *look up fact* was also a common intention (62%).

The users who had the intention of *exploring* the site used the most common navigation strategy, *topical search in a search engine*, in only 15% of the cases. On the other hand using a *search engine to find a known site* and *direct navigation by typing in the URL* were much more frequent (χ^2=11.396, df=4, p<0.05). The "*learner*" frequently (47%) arrived by *a topical search engine search* (χ^2=10.759, df=4, p<0.05). *Direct navigation* was used in large extent, 22% via a *bookmark* and 19% by *typing in the URL*, when the intention was to *look up some fact* (χ^2=25.910, df=4, p<0.01). The users who had the intention of *exploring* the site was statistically younger than the average. Otherwise there were no statistically significant relationships between the intentions and age group, gender or level of search skills.

Between two of the intentions there was statistically significant relationships with the task context in KID. First, *look up fact* was related to the *work* context (χ^2=10.666, df=4, p<0.05). Second, *curious* was related to *by coincidence* (χ^2=52.454, df=4, p<0.01). There were no significant relations between intention and gender or age group in the survey.

9.5 Comparison of log and survey data

To enable the comparison of navigation strategies the five navigation strategies from the survey was grouped into the three basic strategies from the log analysis: direct, link and search engine (Levene, 2010). The data is presented in Appendix 24. In the dataset concerning KID navigation by search engine was the dominant navigation strategy with 53% and 56% respectively in both datasets (Figure 9.4). For the other two resources the similarity between the datasets was low. Perhaps is the distribution of the navigation strategies is skewed because not all users saw the invitation on the front page. In the surveys for both ADL and Poma it looks like direct navigation is overrepresented, at least compared to the log data. Alternatively, the users arriving through direct navigation might be more prone to answer the web survey. In all three cases respondents using direct navigation are overrepresented in comparison with the frequency found in the log files, especially in ADL and Poma, where the difference was over 40 percent. Perhaps this is due to prior knowledge about the resource.

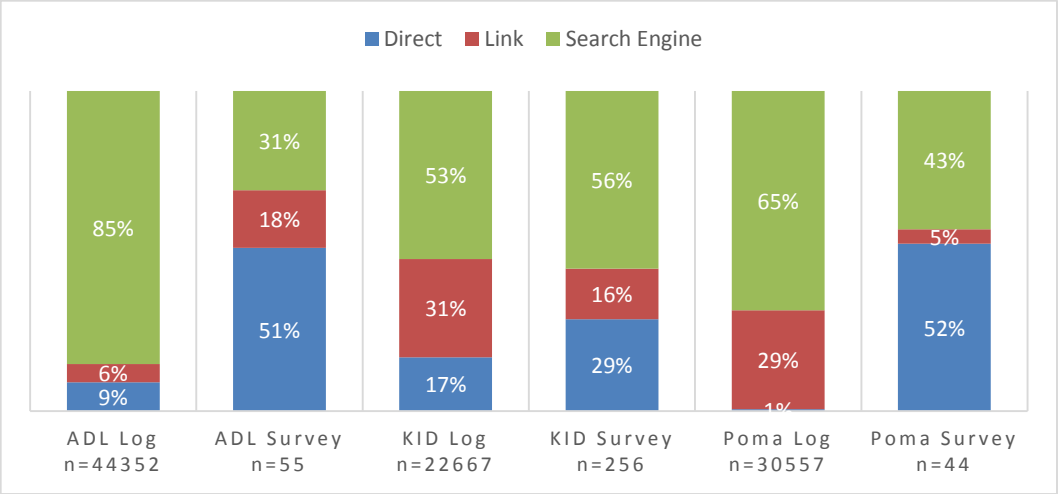


Figure 9.4. Comparison of the distribution of navigation strategies in logs and surveys.

The corresponding distributions of the logs and the survey can also be observed in the users' country of origin (Figure 9.5), which are much more similar than the navigation strategies (Figure 9.4). The logs cover the period from October to December 2010 and the surveys were answered in January-February 2012.

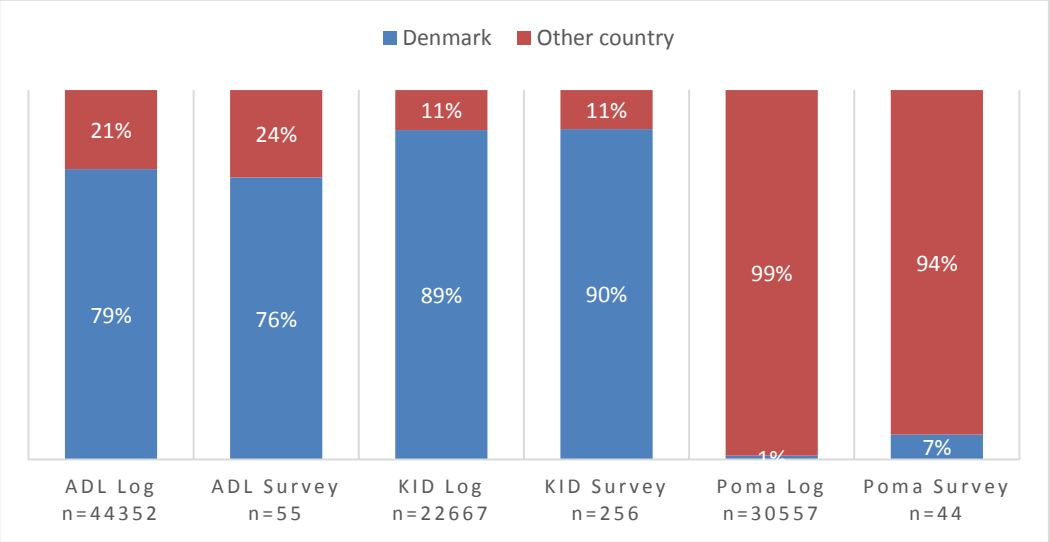


Figure 9.5. Comparison of the distribution of country of origin in logs and surveys.

The distributions of the navigation strategies and the country of origin, and how the great similarity between the two datasets can be looked at are discussed in the conclusion.

9.6 Chapter summary

The survey has provided a complementary description of the users and their usage to the findings in the log analysis. The answers to the two addressed sub-research questions: *Which demographics characterize the users (RQ2e)?* and *Why do users visit the resources (RQ2f)?* In addition the participants were also asked about their navigation to the resource. The survey participants used other navigation strategies than search engine navigation to a much larger degree than the users in the logs.

Which demographics characterize the users? According to the survey results the users are of mixed ages and they are evenly distributed between genders. The respondents of the ADL and KID surveys are primarily from Denmark (more than 70%), whereas the Poma respondents are from other parts of the world. This is confirmed by the findings in the logs, where the Poma users are primarily from Spanish speaking countries.

The question *Why do users visit the resources?* is answered by the survey questions about intention with and context of the visit. In the survey answers from KID there were two distinguishable groups of users. Younger females visited the resource in a work context and older men visited KID for hobby or leisure reasons. Probably the same user groups are present in the other resources, and in Poma there is a third group, the users' study context. Learning is the most frequent intention in all three resources, followed by exploring and looking up fact.

As expexted, the cultural heritage resources are accessed and used by different groups of users and for different reasons. The use of cultural heritage on the web might be similar to the usage of public libraries, diversified and complex. There are a great difference between Poma and the other two resources. Poma is smaller and has no I-level, no information about the digitized objects in a condensed form which easy to access. It has another scope than ADL and KID as a historical document which answers more to topical questions (see Chapter 8). Poma is primarily used in another context, *school and study*, not *hobby or leisure* or *work* as in ADL and KID. Different contexts might give different user behavior.

10 Discussion

In this chapter the research questions will be answered and discussed. The limitations of the research will be addressed. The overall purpose of the present study aims to map and analyse people's use of the digitized cultural heritage and digital cultural resources in everyday life, as stated in the introduction. But it also aims to relate the actions of the users to the actual information environment. ELIS is used as an analytic framework for interpretation of the findings in an everyday context, and the IS&R framework serves as a conceptual foundation leading to my own conceptual framework, the URI model in Figure 2.11. To capture the real usage of cultural heritage resources the main method has been the analysis of log files. The approach is explorative because both the conceptual framework and the findability analysis of CH web resources are developed for the present study, and no previous research is comparable.

The research project is neither pure quantitative nor pure qualitative, it walks the line between the two research approaches. Aspects of both quantitative and qualitative research are combined into a mixed methods approach. The quantitative methods have elements of interpretation built-in and the qualitative methods are used to represent or quantify interpreted aspects or features. Data derived from large datasets, like log files from a web server needs to be critically questioned. There is no exact way of measuring user behaviour on the web and all concepts and measures used in the web industry needs to be reflected upon.

The mixed methods approach, as deployed in the present study with a triangulation of data sources, generates a number of snapshots of the resources, the users and their usage. The URI model (Figure 2.11 and Figure 5.1) illustrates the relative importance of the different aspects in the study, and in Figure 5.3 the methods form a methodological triangulation. In Figure 10.1 the triangulation is displayed together with the applied indicators for respective methods.

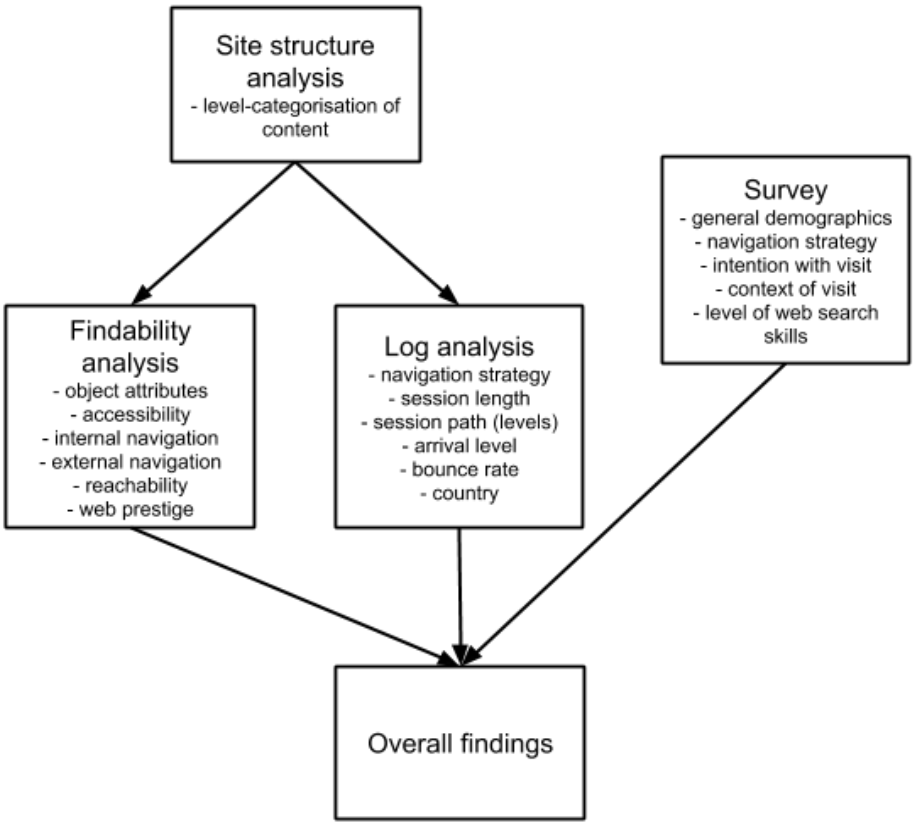


Figure 10.1. The indicators placed under the different methods (based on Figure 5.3).

Based on the answers to all the sub-research questions in the corresponding chapter summaries the two main research questions will be answered and discussed in sequential order; first the findability aspects in RQ1 and then the usage aspects in RQ2. The limitations of the research will be discussed before the general discussion concerning the research results.

10.1 How findable are the heritage resources and their objects?

Findability as a concept was first addressed in order to answer the first research question (RQ1). Six aspects of web findability were identified as central based on previous research combined, with literature from the professional fields of web design and search engine optimization: *attributes of the object*; *accessibility*; *internal navigation*; *internal search*; *reachability*; and *web prestige*. The first two, object attributes and accessibility, are characteristics of the single object, whereas internal navigation and internal search are site or resource dependent features which have implications on all objects within the resource. Reachability is also a structural feature of the resource, but it is one of two aspects concerning the web presence. The other aspect is web prestige, which is the only aspect that is based on external actors. The aspects important for finding objects within a resource are combined into the concept of internal findability. In the same

way, external findability is a combination of the aspects crucial for an object to be findable from outside the resource from the web. Seven indicators were chosen for evaluating the six aspects, one for each aspect except for object attributes, which was measured using two indicators (in Chapter 3). A representative and typical set of objects for each resource were evaluated with the framework, as an illustration of the findability of the resource (in Chapter 7).

The answer to research question one is that the studied resources and their objects in general are findable. The findability could be improved, but there are no serious issues with the findability, besides the lack of PageRank for the objects in KID. KID uses the secure protocol https instead of the normal http protocol and therefore the KID-objects do not have any PageRank-values. The objects are indexed by Google but when the PageRank is zero the objects are ranked low in the results page, well below similar objects with higher PageRank.

It is hard to discuss the findability of one resource compared to other resources based on the findability analysis without a complete findability evaluation of all the comparable resources. The three studied resources are quite large and professionally developed. It might be suspected that smaller resources suffer from more findability issues as they might be produced with a more limited budget. The resources studied were primarily chosen for usage reasons (see Section 3.4.4), and not for the findability analysis. If the findability framework had been the only focus of the study other cultural heritage resources could have been chosen for more extreme results in the findability analysis.

The findability framework as expressed in Table 5.8 is a major outcome of the thesis work and an answer to RQ1. In the findability framework many possibilities exist for adjusting the weighting, scoring and even the measured aspects. The framework could alse be used for different purposes, for example for performance analysis and comparison. The flexibility of the framework and the fact that the investigators must choose a priori the scores in the approach are both a strength and a weakness. It is at the present time impossible to use the framework in more than an informed and openminded manner because of the lack of research concerning findability. The development and use of the findability framework is a part of the exploratory approach taken in the thesis.

Other alternatives of the findability framework have also been discussed, e.g. awarding points to the graded findability measures in a completely other way (see Section 7.5). Another possibility is the change of weighting between the different aspects within the framework to highlight weaknesses in the resource in relation to different navigation strategies (see Section 7.6). The potential of automating the whole process of findability evaluation for studying findability in a large scale was discussed in Section 7.7.

10.2 How do users find and use the cultural heritage resources?

The second research question and the sub-research questions concern the users and the usage of the CH resources. The first sub-research question concerns the navigation strategies: *Which navigation strategies are used by the users to access the resources (RQ2a)?* The users most often navigate to the resources through a web search engine. Direct navigation and navigation by links are used in between 5% and 30% of the session. The second sub-research question addresses the question of where in the resources the users arrives: *On what level in the resources do the users arrive (RQ2b)?* Generally the users arrives at all levels. They arrive often at the navigational top level by direct navigation due to the nature of the navigation strategy. Users navigating with search engines often arrive at the informational level because of the matching between their query and the content of the page.

The internal navigation behaviour in the CH resources is in focus in the third sub-research question: *How do they navigate within the resources (RQ2c)?* In terms of navigation the users uses the resource in different ways. A large group of users only visits one page; they "bounce" away when the first object has been viewed. In the other end of the spectrum a minority of the users visit all levels in the resource within a single session. Cultural heritage objects are viewed in a third of the sessions in ADL, in half of the KID sessions, and in two thirds of the Poma sessions. To answer the sub-research questions the 15 path types have been developed (see Section 8.2.3). The path types may be seen as information pathways (Section 2.2) and some have berrypicking traits (Bates, 1989). The next sub-research question is connected to the previous as it also concerns internal navigation: *How many objects do the users access in a session (RQ2d)?* The average session length varies between 3.7 and 9.0 object views per session. But a large share of users only visits one page, as mentioned above.

The last two sub-research questions were addressed in the web survey and they concerns the questions of why the CH resources are visited and by whom: *Why do users visit the resources (RQ2f)?* Based on the survey findings the resources are most often used for learning activities. The learning occurs in different contexts: hobby or leisure, work, or school or study. The answers in the survey also stress that the CH resources are visited for both hedonic and utilitarian purposes and the intentions with the visits are both to look up things and to explore. And the las sub-research question: *Which demographics characterize the users (RQ2e)?* The users of the ADL and KID are mainly from Denmark and the Poma users are mainly from Spanish speaking countries. The other demographic properties differs greatly. Two typical groups of users were identified in the survey answers from the KID respondents: younger, female professional visitors and older hobbyists, most often males.

The partial results of the sub-research questions contribute to answering the overall RQ2: *How do users find and use the cultural heritage resources?* And the answer relies heavily on how they are formulated. The empirical findings and the answers to the sub-research questions have shown, that the cultural heritage on the web is found in different ways and for different reasons in different

contexts. The navigation to and within the resources are diverse, despite some revealed patterns. Human behaviour is hard to capture with a handful of indicators, but some general trends have been identified, e.g. the large reliance of Google web search and the large use of the information about the cultural heritage objects and its creators in everyday life. One conclusion is that the CH resources are searched and used for different types of searches, e.g. for look-up searches and for exploratory searches (Section 4.2.1).

As can be seen in Appendix 21 it is possible to investigate the relation between session path types and navigations strategies more closely. Another possibility is to explore single sessions in-depth, for example the connection between the queries used in the referring search engine and how the session evolves within the cultural heritage resource and which objects the user visits. Common characteristics for sessions of different length could be explored, for instance the question: Are there similarities between all short or very long sessions?

10.3 How can the different datasets be analysed together?

When working with different types of data the question is to what extent it is possible to analyse the data together. When comparing the distribution of the navigation strategies and the country of origin between the log files and the survey answers great similarities but also differences were found. In the case of KID both the navigation strategy distribution and the country of origin distribution are similar and then it is tempting to draw too many conclusions based on the two datasets combined.

Which conclusions can be made when combining survey data with analysis of logs? When combining different methods there will always be problems in handling the datasets together. In the present study one drawback is that the logs and the survey cover different time periods. There is no evidence that the survey participants are present in the log files and vice-versa. But both datasets contains real users with their own, genuine intentions and tasks which have led them to the site. It is important to remember that the log files contains all real usage of the covered time period. If the survey was distributed during the same time as the logs were collected, the pop-up survey might have interfered with the usage of the site. The central issue is about the representativeness of the survey. As the survey is based on a convenience sample, all participants volunteered, it is impossible to say how representative the sample is. As a user it is very easy to skip an invitation to survey in a pop-up window on the web. But at the same time the survey may be seen as somewhat representative despite it is not random or overlap in time with the log files. Both the distribution of the navigation strategies and the country of origin correspond to a large degree in the two datasets. Based on this observation some general descriptions of the users and their usage have been drawn.

In theory the conceptual framework supports a combined analysis of usage and findability data together. But in the present study with large and abstract datasets covering the usage and the

findability data derived from an evaluation according to specified criteria, the nature of the two datasets is too different. When relating the usage data with the findability it is only possible to discuss the impact of specific features of the site on findability or weaknesses found in the findability analysis influencing the usage patterns. In a study with a more controlled environment, e.g. in laboratory settings, the two kinds of data might be possible to correlate more conclusively. If the findability could be done on all objects in a resource and the usage data could be analysed in relation to specific objects, then the datasets perhaps could be analysed together. Or if the usage data was more qualitative, or at least just covered a specific user group (e.g. by client logging) and the findability analysis covered a larger number of indicators, then the framework might support a combined analysis.

In the present study the framework strengthens the use of mixed methods and the results are a number of indicators, measures and patterns of the usage, the users, and the findability and structure of the resources. These findings cannot be combined in any representative way, but they give valuable insights on the usage of the studied cultural heritage resources and on cultural heritage resources in general.

The mixed methods research design with triangulation does not overcome all issues when working with different kinds of data. But the research design in combination with the conceptual framework makes it possible to begin discussing for example usage in relation to findability.

In the model in Figure 10.2 the three webometric levels are combined with the object model (Figure 2.7b). User actions in form of web navigation strategies are at the usage level and the degree of findability is a part of the structural level. The content at the content level is given by the resources studied. The actions at the usage level are to some parts driven by the content and structure levels, and to some parts by the users' intentions and information need. Each heritage object is contained within a resource (website, database) and is therefore made available on the web. Users have to find their way through both the web and the resource to get to the object.

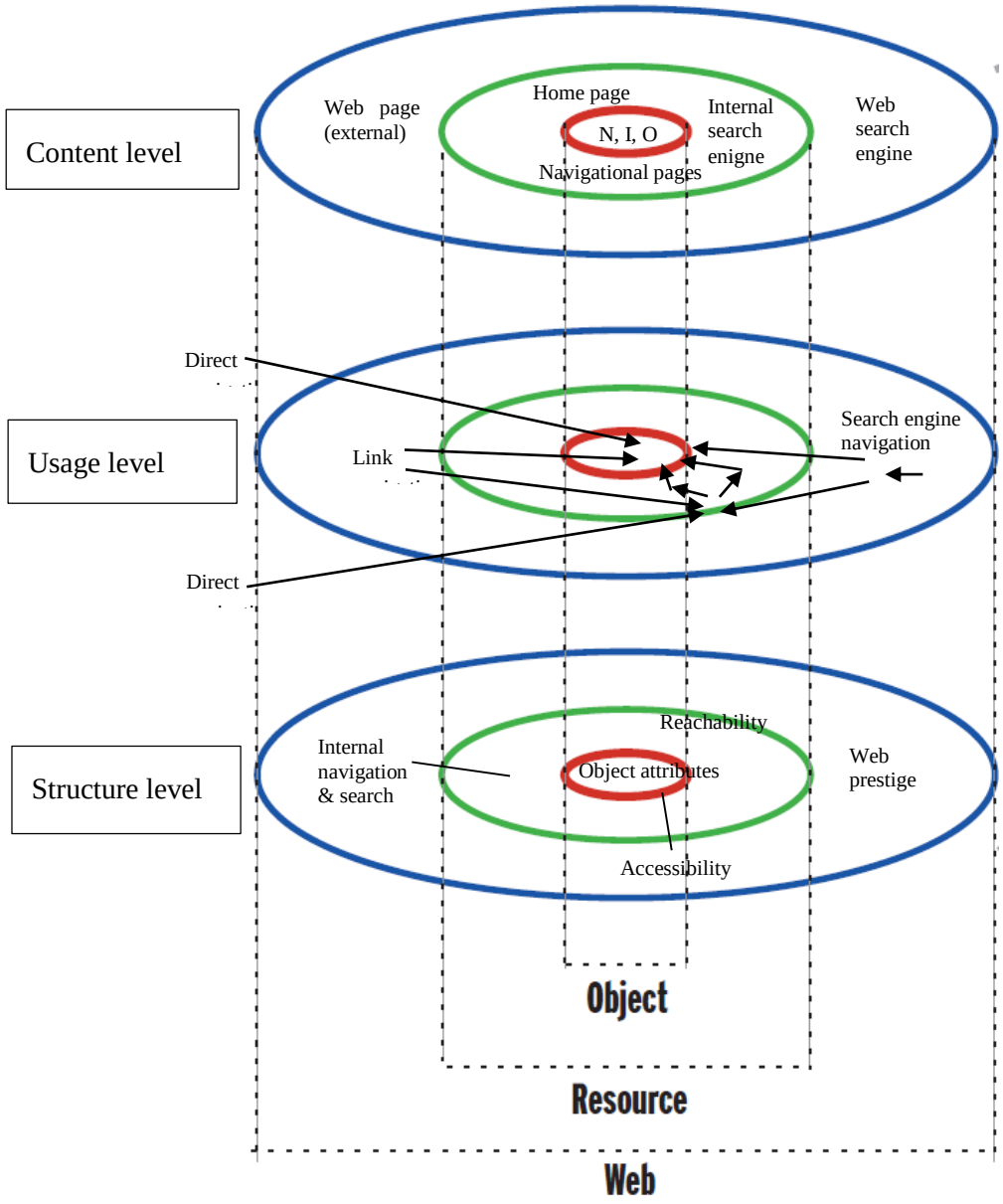


Figure 10.2. The three webometric levels in the object model (Figure 2.7b), where the actions at the usage level depends on all the input values into the information search process as illustrated in Figure 2.11, at a specific moment in time.

A number of potential relations between the levels can be discussed based on the URI model and Figure 10.2. The relatively low use of search engine navigation in KID (Figure 8.1) might depend on the relatively low external findability due to the lack of PageRank (Table 7.5). The search engine users finds objects from other resources instead, i.e. objects ranked higher in the search engine results page. Another possible explanation is a high specificity of the SAPs in KID, which might not match the queries of the users' (the queries in Appendix 16 are queries that have led users to the resource).

The heavy use of the informational objects in ADL and KID probably depends on the correspondence to the information needs of the users. But the usage may also depend on the object attributes of the different kinds of objects. The CH objects may have more specific metadata or just be full text, whereas the Informational objects may contain more general terms which also are used in the queries of the users.

Search engine navigation is the most frequently used navigation strategy, probably because of the general use of web search engines, but it may also indicate high external findability of the objects in the resources.The objects in the CH resources are placed in the top of the search engine results page because they are the most relevant objects available.

The focus on I and O objects in the internal search engines in the resources may result in a limitation of the users' navigation behaviour with the resources. Maybe this is not a problem as the goal of the resources is to lead the users to the digitized cultural heritage objects. At the same time the conception of how web search works is dominated by the web search engines, where all kinds of objects are findable.

Besides relating different empirical findings on different levels, Figure 10.2 (and the URI model in Figure 2.11) can be used to formulate research questions for future research. For example to address the potential relations discussed above, which have to be investigated further.

10.4 Limitations of the research

As presented in chapter one some external constraints to the research project presented themselves which had impact on the research design (see Section 1.7). The study covers a broad area of research, including parts from several research fields such as interactive information retrieval, human information behaviour, ELIS and webometrics. The transdisciplinary approach together with the mixed methods research design generates new perspectives on – and new descriptions of – the usage of web resources. But this also means that each research area and each method is not used to their full potential. The lack of depth is the price paid for doing explorative multi-method research, but is compensated with the new perspectives and the development of new concepts and methods (e.g. the site structure analysis and the findability framework and their indicators).

The first major limitation is that only three cultural heritage resources are studied, and no audio or video collections were included. Log analysis as a method imposes some limitations on the research due to the nature of the data in the logs. By means of the method only questions about how users behave, not why, are possible to answer.

The web survey has given valuable insights about the users and why they engage in interaction with the information. But both in the case of ADL and in Poma the samples were small (~50 participants). And the survey findings are very hard to connect to the findings in the logs. The

fact that the logs and the survey cover different time periods is an issue, but is not very important because the respondents would not have been statistically representative in any case.

The site structure analysis is based on the present framework and the outcome might have been different with other research questions or if other types of resources were analysed. The findability analysis is well anchored in the current practices of web design and search engine optimisation, as well as in web research. But it is the first attempt to evaluate web findability on an aggregated level. Alternative weighting and scores of the indicators in both external and internal findability are possible and illustrated briefly, sections 7.5-7.6.

The research has become more qualitative during the research process than envisioned as there are interpretive elements, e.g. the site structure analysis and the implementation of the findability analysis. Neither the log analysis nor the surveys are based on statistically representative samples, so it is not possible to do statistical generalisations based on the results. But as discussed in section 5.6 it is possible to discuss the existence of patterns in both the behaviour of the users and the findability. There are probably more common patterns to be found between cultural heritage resources online than there are unique patterns in individual resources.

10.5 The results in light of ELIS and IS&R

The focus of the thesis is the general public's use of digitalized cultural heritage resources (see Section 1.1). The empirical results are placed in a everyday perspective, i.e. the ELIS framework (see Section 2.2). In ELIS several specifc concepts are defined and used: everyday information practice, information source horizons, zones of source preference, and information pathways. The basis in ELIS is the user's interests, to quote Savolainen: "The objects in the everyday world capture a person's attention through his or her *interests*." (Savolainen, 2008, p. 56). Interests are also the foundation in the Serious Leisure Perspective (Stebbins, 2007) which addresses different types of leisure projects. The question about interest is hard to address directly. In the survey two questions concerns interests indirectly: the reason of the visit and the context of the visit. Being curious as a reason for visiting the CH resource. This indicates an interest, but other possible answers may also include interest indirectly (answer a question, exploring, learning, looking up a fact, or other). The question about the context of the visit has one clear answer connected to interest, the hobby or leisure context. But the other answer alternatives may also contain pure interest visits (work, by coincidence, school or study, or other). In this sense it is impossible to discuss ELIS information practices. The usage data cover all types of information practices regardless of context or motivation of the visit.

The log data and the survey answers give indications on the information source horizons, zones of source preference and the information pathways. The queries in the referring search engines are to a large extent navigational: the users know where they want to go on the web and use the search engine as a shortcut. Many of the survey respondents knew the URL, wrote it in the

browser or used a saved bookmark, to access the CH resource. This indicates that the CH resources are placed in a central zone in many of the users' information source horizons. The information pathways of the users are partially visible when studying the referring sites (search engines excluded). Wikipedia is an important referrer for the CH resources, and not simply the Danish, English and Spanish versions but other language versions as well. Other central web resources on the topic are also important resouces earlier on the information pathways of the users. With one exception social media is absent as an important stop before arriving at the CH resources. Supported by the empirical findings the CH resource might be said to be quite well known by a considerable group of everyday life users. How large or small this group is in relation the whole potential number of users in the general public is impossible to say.

The IS&R framework (Ingwersen & Järvelin, 2005) has been an inspiration and is used as a foundation for the conceptual framework (Figure 2.8). As the original IS&R framework does not take the complexity into account when the users moves between the web and specific resources and within resources, a modified version of the model was developed (Figure 2.9) where it is possible to make a distinction between the two types of navigation. The web IS&R model also includes the object model from Figure 2.7b, where objects are seen as embedded in a resource on the web. With the URI model in Figure 2.11 I have made a model that explicitly builds on the IS&R model, and which specifies the interaction s between user and resource. IS&R framework has thus been developed to include web IR.

11 Conclusions

In this chapter overall conclusions are drawn. Contributions to research in Library and information science are presented, as well as concluding remarks and directions for future research.

11.1 Overall conclusions

The main conclusion of the thesis is that the users search for and visits pages containing information about the digitized cultural heritage objects and the creators of the objects rather than the digitized CH objects themselves. This meta information is often looked upon as a byproduct or a context to the cultural heritage that is digitized and made available. Perhaps the CH institutions should prioritize the creation and compilation of the objects about the CH to match the information needs of the everyday user. The CH objects themselves are often available in different formats, e.g. at the museum or at the local library. CH objects are often unique and making them available online is important, like the Poma chronicle which is an extraordinary resource of international importance. But as the empirical results has shown the information about the cultural heritage sought for is very important. This means that the digital efforts of the CH institutions should be closer to Wikipedia than to for example Europeana, the European search portal for CH objects. The subject expertise in the CH institutions can play a new role to match the information needs of the general public. This is a public which uses the CH resources in different contexts, for study or work, hobbies or leisure.

The findability of the CH resources are generally good in the analysed resources. The studied resources are large in view of the amount of content. All three are published by large institutions, and they may not be representative for smaller CH resources. The degree of findability could be increased for all three resources, importantly the number of SAPs can increased to improve findability. Achieving and maintain good findability is an ongoing endeavour.

11.2 Contributions to research

The URI model (Figure 2.11) with the distinction of the interaction between the user and the system into three dimensions, content, usage, and structure, is a major conceptual contribution. Previous research has explored the importance of domain knowledge versus search knowledge during information interaction, but with the URI model it is possible to do the corresponding analysis of the system side during information searching. This is because the content level is seen

as query-dependent and the structure level is seen as query-independent. The model may be useful on clarifying issues where the dimensions sometimes are mixed up, for example relevance ranking algorithms (query-dependent) and the PageRank algorithm (query-independent). On a more general level the model can be useful to research in both Interactive Information Retrieval (IIR) and Human Information Behaviour (HIB).

The development of the IS&R model for web use, at least in the context of navigation and search, might increase the usefulness of the IS&R framework as there is more research on web search than on the use of specific information systems, at least from an information searching perspective.

The analysis of the log files has shown that some concepts, often used as indicators, are hard to interpret in an unambiguous way. For example the bounce rate can have totally different meanings depending on the investigator's perspective. The present research calls for a careful use or interpretation of web analytic concepts like bounce rate, returning visitors, etc., both in research and practice. The analysis of the logs and the surveys has increased the knowledge about the usage of cultural heritage resources as well as about the users.

The whole findability framework developed in this thesis is a significant contribution to research. Some aspects of findability, as it is defined here, have been studied before as search engine visibility or accessibility, but the findability framework introduced in this thesis is significantly more comprehensive and coherent than any previously presented. The framework could be developed further in several directions, that go beyond the specific setting applied in this thesis. The framework can be used for performance evaluation and the points awarded for the aspects can be logarithmic rather than linear (Section 7.5). As illustrated in Section 7.6 it is possible to change the weighting of the aspects to highlight different matters. The analysis could be automated and applied on whole sites or even several sites for comparison, as discussed in Section 7.7. Another possibility is to add more indicators for each aspect and thereby evaluate the findability in a more complex and richer way. But first, as stated elsewhere, the findability framework needs to be tested on other kinds of web resources.

11.3 Implications for practice

The web findability framework can be used by site owners and designers, as an analysis method for pinpointing the weak spots on their web resources with respect to web navigation and findability.

The navigation strategies used by the visitors give important clues both as to the users and the resource. Direct navigation indicates regular, highly motivated returning users, and it can be of interest to study which objects or types of objects they return to. A large share of link navigators might depend on one or several central referring sites which generates a lot of traffic, but it might also indicate that it is hard for the users to find the resource and its objects in the web search

engines. The link navigation share is high because the search engine navigation share is low. In all types of resources the search engine navigation ought to have a large share of the total number of navigation strategies. The query analysis of the terms used in the referring search engines points to the importance of indexable text to be found by the users through the search engines. The text on any page should be taken serious and be of a minimum size, and without metadata other type of objects, like pictures, easily becomes unfindable. Content production and indexing of objects is time consuming, maybe the possibilities to automatically add metadata should be explored by cultural heritage resource owners. Another possiblilty is crowd sourcing, i.e. to let the users contribute with keywords.

Generally all cultural heritage resources ought to have a findability plan. A plan covering the structural and technical aspects based in the design and construction of the resource, as well as the on-going effort to increase the web prestige and improve the contents, e.g. adding metadata and text. On the web there is a constant struggle for web prestige in competition with other resources. As argued by Walter (2008) findability is not a specific function, but a goal for all involved actors to fight for (see Section 3.2.1). Search engine optimisation might be an important aspect to work with and it is easy to outsource, but more important is the quality of the contents in the resource. To drive a lot of traffic to the resource is meaningless if the large majority bounces because the resource does not match the expectations of the visitors.

11.4 Concluding remarks and directions for future research

In general the present study shows the importance of complementing log analysis with other methods of data collection, but it is not obvious how it should or could be done in the research literature. Contextual information about the users and their usage is central to interactive IR. Measures and methods for studying and evaluating information seeking and retrieval in context has to be further developed and different mixed method approaches has to be tried and tested both in laboratory and real world settings.

The result raises new questions. Do users do any distinction between cultural heritage resources from other resources on the Web? Does the digitalized cultural heritage match the users' needs for information and experiences? Perhaps information about the cultural heritage is wanted more than actual digitalised objects? But content creation demands the cultural heritage institution to engage other professions or redefine the work descriptions of the experts in the institutions. And if information about the cultural heritage should be produced, where should it be published? On specific web sites which might require some serendipity to find or on a well-known, well trafficked public site like Wikipedia? Or perhaps both by parallel publishing it and get traffic to the institutional site from Wikipedia.

Research about how citizens perceive the cultural heritage, both online and offline, is needed. But most important, there is a great need of research connecting the behaviour of the users, or other

actors, with the attributes of the information environment. The information environment includes both digital and physical spaces not just single IR-systems, a largely unexplored factor in IS&R research.

The two traditions within IS&R, Information Seeking (or Human Information Behaviour) and Interactive Information Retrieval are hard to combine. The present study has been an attempt to carry out IS&R research, but there is a fundamental difference between the two traditions. According to Fidel the HIB research results and models are mainly descriptive. And so are the results on the usage of the cultural heritage resources in the present study. In IIR on the other hand the models are normative, like the findability analysis. The two types of research, descriptive and normative are hard to combine (Fidel, 2012). And results on human information behaviour will still be mainly descriptive, even in the future, but the findings will be of interest – the exploration is on-going as new information environments continuously are invented and new "users" are born.

References

ACRL. Information Literacy Competency Standards for Higher Education Retrieved 24/10, 2013, from http://www.ala.org/acrl/standards/informationliteracycompetency

Alfort, E. (2013). *The expression of a need : understanding search*. Copenhagen: PhD School of Economics and Management, Copenhagen Business School.

Almind, T. C., & Ingwersen, P. (1997). Informetric analyses on the World Wide Web: methodological approaches to 'Webometrics'. *Journal of Documentation, 53*(4), 404-426. doi: 10.1108/EUM0000000007205

Alpert, J., & Hajaj, N. (2008). We knew the web was big Retrieved 22/07, 2013, from http://googleblog.blogspot.se/2008/07/we-knew-web-was-big.html

Angelov, I., Menon, S., & Douma, M. (2010). Finding Information: Factors that Improve Online Experiences. In H. H. Yang & S. C.-Y. Yuen (Eds.), *Handbook of Research on Practices and Outcomes in E-Learning: Issues and Trends* (pp. 493-506). Hersey, PA: IGI.

Azzopardi, L., & Bache, R. (Eds.). (2010). *On the relationship between effectiveness and accessibility*: ACM. doi: 10.1145/1835449.1835667

Azzopardi, L., & Vinay, V. (2008). *Retrievability: an evaluation measure for higher order information access tasks.* Paper presented at the Proceeding of the 17th ACM conference on Information and knowledge management.

Bar-Ilan, J. (2004). The use of web search engines in information science research. *Annual Review of Information Science & Technology, 38*(1), 231-288.

Bashir, S., & Rauber, A. (2009). Analyzing document retrievability in patent retrieval settings *Database and Expert Systems Applications. 20th International Conference, DEXA 2009, Linz, Austria, August 31 – September 4, 2009. Proceedings.* (pp. 753-760): Springer. doi: 10.1007/978-3-642-03573-9_63

Bates, M. (1979). Information search tactics. *Journal of the American Society for Information Science, 30*(4), 205-214. doi: 10.1002/asi.4630300406

Bates, M. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review, 13*(5), 407-424. doi: 10.1108/eb024320

Bates, M. (1990). Where Should The Person Stop And The Information Search Interface Start? *Information Processing and Management, 26*(5), 575-591. doi: 10.1016/0306-4573(90)90103-9

Bates, M. (2002). Toward an integrated model of information seeking and searching. *The New Review of Information Behaviour Research, 3*, 1-15.

Bates, M. (2009a). Information *Encyclopedia of Library and Information Sciences, Third Edition* (pp. 2347-2360): Taylor & Francis. doi: 10.1081/e-elis3-120045519

Bates, M. (2009b). Information Behavior *Encyclopedia of Library and Information Sciences, Third Edition* (pp. 2381-2391): Taylor & Francis. doi: 10.1081/e-elis3-120043263

Bates, M., Wilde, D. N., & Siegfried, S. (1993). An analysis of search terminology used by humanities scholars: the Getty Online Searching Project Report Number 1. *The Library Quarterly, 63*(1), 1-39.

The behaviour/practice debate: a discussion prompted by Tom Wilson's review of Reijo Savolainen's Everyday information practices: a social phenomenological perspective. (2009). *Information Research, 14*(2).

Belkin, N. J. (1996). *Intelligent Information Retrieval: Whose Intelligence?* Paper presented at the Harausforderungen an die Informationswirtschaft. Informationsverdichtung, Informationsbewertung und Datenvisualisierung, Proceedings of the 5th International Symposium for Information Science (ISI '96).

Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval: Part I. Background and theory. *Journal of Documentation, 38*(2), 61-71. doi: 10.1108/eb026722

Bergman, M. K. (2001). The deep web: Surfacing hidden value. *Journal of Electronic Publishing, 7*(1), 07-01. doi: 10.3998/3336451.0007.104

Bergman, M. M. (2011). The Good, the Bad, and the Ugly in Mixed Methods Research and Design. *Journal of Mixed Methods Research, 5*(4), 271-275. doi: 10.1177/1558689811433236

Berners-Lee, T. (1996). Web Architecture: Generic Resources Retrieved 22/07/2013, 2013, from http://www.w3.org/DesignIssues/Generic.html

Björneborn, L. (2004). *Small-world link structures across an academic web space: a library and information science approach : PhD thesis*. Royal School of Library and Information Science. Copenhagen. Retrieved from http://pure.iva.dk/files/31034741/lennart_bjorneborn_phd.pdf

Björneborn, L. (2008). Serendipity dimensions and users' information behaviour in the physical library interface. *Information Research, 13*(4), 13-14.

Björneborn, L. (2011a). *Behavioural Traces and Indirect User-to-User Mediation in the Participatory Library*. Paper presented at the Information Science and Social Media : Proceedings of the International Conference Information Science and Social Media, ISSOME2011, August 24-26, Åbo/Turku, Finland, Åbo.

Björneborn, L. (2011b). Genre connectivity and genre drift in a web of genres. In A. Mehler, S. Sharoff & M. Santini (Eds.), *Genres on the Web : Computational Models and Empirical Studies* (pp. 255-274): Springer. doi: 10.1007/978-90-481-9178-9_12

Björneborn, L., & Ingwersen, P. (2004). Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology, 55*(14), 1216-1227. doi: 10.1002/asi.20077

Black, T. R. (1999). *Doing quantitative research in the social sciences : an integrated approach to research design, measurement and statistics*. London: SAGE.

Borges, J., & Levene, M. (2007). Evaluating variable-length markov chain models for analysis of user web navigation sessions. *Knowledge and Data Engineering, IEEE Transactions on, 19*(4), 441-452.

Borlund, P. (2000). *Evaluation of interactive information retrieval systems*: Åbo Akademis Förlag.

Boyce, B. R., Meadow, C. T., & Kraft, D. H. (1994). *Measurement in information science*. San Diego, CA: Academic.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems, 30*(1-7), 107-117. doi: 10.1016/s0169-7552(98)00110-x

Broder, A. (2002). A taxonomy of web search. *SIGIR Forum, 36*(2), 3-10. doi: 10.1145/792550.792552

Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., . . . Wiener, J. (2000). Graph structure in the web. *Computer networks, 33*(1), 309-320. doi: 10.1016/S1389-1286(00)00083-9

Bruns, A. (2008). *Blogs, Wikipedia, Second life, and Beyond : from production to produsage*. New York, NY: Peter Lang.

Bryman, A. (1988). *Quantity and quality in social research*. London: Unwin Hyman.

Buckland, M. K. (1991). Information as thing. *Journal of the American Society for Information Science, 42*(5), 351-360.

Buckland, M. K. (1997). What Is a ``Document"? *Journal of the American Association for Information Science, 48*(9), 804-809.

Byström, K., & Hansen, P. (2005). Conceptual framework for tasks in information studies. *Journal of the American Society for Information Science and Technology, 56*(10), 1050-1061. doi: 10.1002/asi.20197

Byström, K., & Järvelin, K. (1995). Task complexity affects information seeking and use. *Information Processing & Management, 31*(2), 191-213. doi: 10.1016/0306-4573(95)80035-r

Canter, D., Rivers, R., & Storrs, G. (1985). Characterizing user navigation through complex data structures. *Behaviour & Information Technology, 4*(2), 93-102. doi: 10.1080/01449298508901791

Case, D. O. (2007). *Looking for information : a survey of research on information seeking, needs, and behavior*. Amsterdam: Elsevier/Academic Press.

Chevillotte, S. (2010). Information Literacy *Encyclopedia of Library and Information Sciences* (Third Edition ed., pp. 2421-2428). London: Taylor & Francis. doi: 10.1081/E-ELIS3-120043727

Chu, H. (2010). *Information Representation And Retrieval In The Digital Age*. Medford, NY: Information Today.

Clark, D. (2011). M3.1.2 - Europeana Log Analysis Report 1: EuropeanaConnect.

Cooley, R., Mobasher, B., & Srivastava, J. (1999). Data preparation for mining world wide web browsing patterns. *Knowledge and Information Systems, 1*(1), 5-32. doi: 10.1007/BF03325089

Croft, W. B., Metzler, D., & Strohman, T. (2010). *Search engines : information retrieval in practice*. Boston, MA: Addison-Wesley.

Denscombe, M. (2010). *Good research guide : for small-scale social research projects*. Buckingham: Open University.

Denzin, N. K. (1970). *The research act; a theoretical introduction to sociological methods*. Chicago, IL: Aldine.

Denzin, N. K. (2012). Triangulation 2.0*. *Journal of Mixed Methods Research, 6*(2), 80-88. doi: 10.1177/1558689812437186

Dervin, B. (1998). Sense-making theory and practice: an overview of user interests in knowledge seeking and use. *Journal of Knowledge Management, 2*(2), 36-46. doi: 10.1108/13673279810249369

Dieberger, A., Dourish, P., Höök, K., Resnick, P., & Wexelblat, A. (2000). Social navigation: techniques for building more usable systems. *Interactions, 7*(6), 36-45. doi: 10.1145/352580.352587

Ding, W., & Lin, X. (2010). *Information architecture : the design and integration of information spaces*. San Rafael, CA: Morgan & Claypool.

Dourish, P. (2004). *Where the action is : the foundations of embodied interaction*. Cambridge, MA: MIT Press.

Ellis, D. (1993). Modeling the information-seeking patterns of academic researchers: A grounded theory approach. *Library Quarterly, 63*(4), 469-486.

Elsweiler, D., Wilson, M. L., & Lunn, B. K. (2011). Understanding casual-leisure information behaviour. In A. Spink & J. Heinström (Eds.), *New Directions in Information Behaviour* (pp. 211-241): Emerald. doi: 10.1108/S1876-0562(2011)002011a012

Enge, E. (2009). *The art of SEO*. Sebastopol, CA: O'Reilly.

Fallows, D. (2006). Surfin for Fun Retrieved 03/06, 2013, from http://www.pewinternet.org/Reports/2006/Surfing-for-Fun.aspx

Fidel, R. (1985). Moves in online searching. *Online review, 9*(1), 61-74. doi: 10.1108/eb024176

Fidel, R. (2012). *Human information interaction : an ecological approach to information behavior*. Cambridge, MA: MIT Press.

Ford, N., & Mansourian, Y. (2006). The invisible web: an empirical study of "cognitive invisibility". *Journal of Documentation, 62*(5), 584-596. doi: 10.1108/00220410610688732

Ford, N., Wilson, T. D., Foster, A., Ellis, D., & Spink, A. (2002). Information seeking and mediated searching. Part 4. Cognitive styles in information seeking. *Journal of the American Society for Information Science and Technology, 53*(9), 728-735. doi: 10.1002/asi.10084

Frankfort-Nachmias, C., & Nachmias, D. (2000). *Research methods in the social sciences*. New York, NY: Worth Publishers.

Fransson, J. (2007). *Effektivare informationssökning på webben : en handbok i konsten att söka information*. Ronneby: Hexa förlag.

Fransson, J. (2010). Det danska digitaliserade kulturarvet. *Revy, 33*(6), 8-11.

Fransson, J. (2011). Findability och informationskompetens vid webbnavigation. *Dansk Biblioteksforskning, 7*(2/3), 55-68.

Fransson, J. (2012). *Intention and task context connected with session in a cultural heritage collection*. Paper presented at the Proceedings of the 4th Information Interaction in

Context Symposium, Nijmegen, The Netherlands, August 21-24, 2012. doi: 10.1145/2362724.2362750

Fricker, R. D. (2008). Sampling methods for web and e-mail surveys. In N. Fielding, R. M. Lee & G. Blank (Eds.), *The SAGE handbook of online research methods* (pp. 195-216). Los Angeles; London: SAGE.

Gerjets, P., & Hellenthal-Schorr, T. (2008). Competent information search in the World Wide Web: Development and evaluation of a web training for pupils. *Instructional Support for Enhancing Students' Information Problem Solving Ability, 24*(3), 693-715. doi: 10.1016/j.chb.2007.01.029

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.

Gudiksen, J. (2005). Formidling. En karakteristik af forskellige formidlingsforståelser og deres kommunikationsteoretiske og forskningstypologiske grundlag. *Dansk Biblioteksforskning, 1*(2), 31-40.

Gäde, M., Petras, V., & Stiller, J. (2010). Which Log for which Information? Gathering Multilinguality Data from Different Log File Types *Multilingual and Multimodal Information Access Evaluation : Lecture Notes in Computer Science Volume 7360* (pp. 70-81): Springer.

Haas, S. W., & Grams, E. S. (2000). Readers, authors, and page structure: A discussion of four questions arising from a content analysis of web pages. *Journal of the American Society for Information Science, 51*(2), 181-192. doi: 10.1002/(SICI)1097-4571(2000)51:2<181::AID-ASI9>3.0.CO;2-8

Halavais, A. (2009). *Search engine society*. Cambridge: Polity Press.

Halpin, H., & Presutti, V. (2009). An ontology of resources: Solving the identity crisis *The Semantic Web: Research and Applications* (pp. 521-534): Springer.

Halpin, H., & Presutti, V. (2011). The identity of resources on the web: an ontology for web architecture. *Applied Ontology, 6*(3), 263-293. doi: 10.3233/AO-2011-0095

Hantula, D. A. (2010). The Behavioral Ecology of Human Foraging in an Online Environment: Of Omnivores, Informavores, and Hunter–Gatherers. In N. Kock (Ed.), *Evolutionary psychology and information system research* (pp. 85-99): Springer.

Harper, S., & Yesilada, Y. (2008). Web accessibility and guidelines. In S. Harper & Y. Yesilada (Eds.), *Web Accessibility: A Foundation for Research* (pp. 61-78): Springer.

Hartel, J. (2003). The Serious Leisure Frontier in Library and Information Science: Hobby Domains. *Knowledge Organization, 30*(3/4), 228-238.

Hartel, J. (2009). Leisure and Hobby Information and Its Users *Encyclopedia of Library and Information Sciences, Third Edition* (pp. 3263-3274): Taylor & Francis. doi: 10.1081/e-elis3-120043076

Hartel, J. (2010). Managing documents at home for serious leisure: a case study of the hobby of gourmet cooking. *Journal of Documentation, 66*(6), 847-874. doi: 10.1108/00220411011087841

Heckner, M., Heilemann, M., & Wolff, C. (2009). Personal information management vs. resource sharing: Towards a model of information behaviour in social tagging systems. *Proceedings from the Third International AAAI Conference on Weblogs and Social*

Media (ICWSM 09), from http://www.aaai.org/ocs/index.php/ICWSM/09/paper/viewFile/212/407

Heinström, J. (2003). Five personality dimensions and their influence on information behaviour. *Information Research, 9*(1), 9-1.

Hersh, W. (2009). *Information Retrieval : A Health and Biomedical Perspective*. New York, NY: Springer-Verlag New York Inc.

Hjørland, B., & Nielsen, L. K. (2001). Subject Access Points in Electronic Retrieval. *Annual Review of Information Science and Technology, 35*(Journal Article), 249-298.

Holdgaard, N., & Simonsen, C. (2011). Attitudes towards and conceptions of digital technologies and media in Danish museums. *MedieKultur. Journal of media and communication research, 27*(50), 100-118.

Holt, N. (2012). *Psychology : the science of mind and behaviour*. London: McGraw-Hill Higher Education.

Huberman, B. A. (2001). *The laws of the Web : patterns in the ecology of information*. Cambridge, MA: MIT Press.

Hung, P. W., Johnson, S. B., Kaufman, D. R., Mendonça, E. A., Hung, P., Johnson, S., . . . Mendonça, E. (2008). A Multi-Level Model of Information Seeking in the Clinical Domain. *Journal of Biomedical Informatics, 41*(2), 357-370. doi: 10.1016/j.jbi.2007.09.005

Huntington, P., Nicholas, D., & Warren, D. (2004). Digital visibility and its impact upon online usage: case study of a health Web site. *Libri, 54*(4), 211-220. doi: 10.1515/LIBR.2004.211

Huvila, I. (2009). Ecological framework of information interactions and information infrastructures. *Journal of Information Science, 35*(6), 695-708. doi: 10.1177/0165551509336705

Hyvönen, E. (2012). *Publishing and Using Cultural Heritage Linked Data on the Semantic Web* (Vol. 2): Morgan & Claypool Publishers. doi: 10.2200/s00452ed1v01y201210wbe003

Høgenhaven, T., & Lundberg Andreasen, L. (2011). *Når internettet har magten : om forsvundne og findbare offentlige hjemmesider*. København: Handelshøjskolens Forlag.

Hölscher, C., & Strube, G. (2000). Web search behavior of Internet experts and newbies. *Computer Networks, 33*(1), 337-346. doi: 10.1016/S1389-1286(00)00031-1

IFLA. (2008). Functional requirements for bibliographic record. Final report: International Federation of Library Associations and Institutions (IFLA).

Ingwersen, P. (1992). *Information retrieval interaction*. London: Taylor Graham.

Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. *Journal of Documentation, 52*(1), 3-50. doi: 10.1108/eb026960

Ingwersen, P., & Björneborn, L. (2004). Methodological issues of webometric studies. In H. Moed, W. Glänzel & U. Schmoch (Eds.), *Handbook of Quantitative Science and Technology Research* (pp. 339-369). Dordrecht: Kluwer Academic Publishers.

Ingwersen, P., & Järvelin, K. (2005). *The turn : integration of information seeking and retrieval in context*. Dordrecht: Springer Verlag.

Jansen, B. J. (2009a). The Methodology of Search Log Analysis. In B. J. Jansen, I. Taksai & A. Spink (Eds.), *Handbook of research om web log analysis* (pp. 99-121). Hersey, PA: Information Science Reference.

Jansen, B. J. (2009b). *Understanding user-Web interactions via Web analytics*. San Rafael, CA: Morgan & Claypool Publishers. doi: 10.2200/S00191ED1V01Y200904ICR006

Jansen, B. J., Booth, D. L., & Spink, A. (2008). Determining the informational, navigational, and transactional intent of Web queries. *Information Processing & Management, 44*(3), 1251-1266. doi: 10.1016/j.ipm.2007.07.015

Jansen, B. J., Taksai, I., & Spink, A. (2009). Research and methodological foundations of transaction log analysis. In B. J. Jansen, I. Taksai & A. Spink (Eds.), *Handbook of research om web log analysis* (pp. 1-16). Hersey, PA: IGI Global.

Jensen, B. E. (2008). *Kulturarv: et identitetspolitisk konfliktfelt*. København: Gad.

Johnson, J. D. E., Case, D. O., Andrews, J., Allard, S. L., & Johnson, N. E. (2006). Fields and pathways: Contrasting or complementary views of information seeking. *Information Processing & Management, 42*(2), 569-582. doi: 10.1016/j.ipm.2004.12.001

Joinson, A. N. (2008, 2008). *Looking at, looking up or keeping up with people?: motives and use of facebook.* Paper presented at the Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, New York, NY, USA.

Jones, W. P. (2007). Personal information management. *Annual Review of Information Science and Technology, 41*(1), 453-504. doi: 10.1002/aris.2007.1440410117

Jones, W. P. (2008). *Keeping found things found : the study and practice of personal information management*. Boston, MA: Morgan Kaufmann Publishers.

Jones, W. P. (2009). Personal Information Management (PIM) *Encyclopedia of Library and Information Sciences, Third Edition* (pp. 4137-4147): Taylor & Francis. doi: 10.1081/e-elis3-120043258

Järvelin, K., & Kekäläinen, J. (2002). Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems, 20*(4), 422-446. doi: 10.1145/582415.582418

Kallinikos, J., Aaltonen, A., & Marton, A. (2010). A theory of digital objects. *First Monday, 15*(6-7).

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM), 46*(5), 604-632. doi: 10.1145/324133.324140

Knight, S. A., & Spink, A. (2008). Toward a Web search information behavior model. In M. Zimmer & A. Spink (Eds.), *Web Search*: Springer.

Kofod-Petersen, A., & Aamodt, A. (2003). *Case-based situation assessment in a mobile context-aware system.* Paper presented at the AIMS2003.

Koll, M. (2000). Track 3: information retrieval. *Bulletin of the American Society for Information Science and Technology, 26*(2), 16-18. doi: 10.1002/bult.143

Kopackova, H., Michalek, K., & Cejna, K. (2010). Accessibility and findability of local e-government websites in the Czech Republic. *Universal Access in the Information Society, 9*(1), 51-61. doi: 10.1007/s10209-009-0159-y

Krug, S. (2006). *Don't make me think! : a common sense approach to Web usability*. Berkeley, CA: New Riders Publishers.

Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science, 42*(5), 361-371. doi: 10.1002/(SICI)1097-4571(199106)42:5<361::AID-ASI6>3.0.CO;2-#

Kulturministeriet. (2009). Digitalisering af kulturarven: endelig rapport fra Digitaliseringsudvalget. København: Kulturministeriet.

Kvale, S. (2008). *Doing Interviews*. London: Sage.

Kvale, S., & Brinkmann, S. (2009). *Den kvalitativa forskningsintervjun*. Lund: Studentlitteratur.

Langville, A. N., & Meyer, C. D. (2006). *Google's PageRank and beyond : the science of search engine rankings*. Princeton, NJ: Princeton University Press.

Laplante, A. (2008). *Everyday life music information-seeking behaviour of young adults: an exploratory study.* Dissertation/Thesis, McGill University.

Larsen, B., Ingwersen, P., & Kekäläinen, J. (2006). *The Polyrepresentation Continuum in IR.* Paper presented at the Information interaction in context : International Symposium on Information Interaction in Context, IIiX 2006.

Lewandowski, D. (2006). Query types and search topics of German Web search engine users. *Information Services and Use, 26*(4), 261-269.

Levene, M. (2010). *An introduction to search engines and web navigation*. Hoboken, NJ: John Wiley.

Limberg, L., Hultgren, F., & Jarneving, B. (2002). *Informationssökning och lärande : en forskningsöversikt*. Stockholm: Skolverket.

Limberg, L., Sundin, O., & Talja, S. (2009). Teoretiska perspektiv på informationskompetens. In J. Hedman & A. Lundh (Eds.), *Informationskompetenser: Om lärande i informationspraktiker och informationssökning i lärandepraktiker* (pp. 36-65). Stockholm: Carlsson.

Lund, N. D., Andersen, J., Dam Christensen, H., Johannsen, C. G., & Skouvig, L. (2009). *Digital formidling af kulturarv : fra samling til sampling*. København: Multivers.

Lutze, H. (2009). *The findability formula : the easy, non-technical approach to search engine marketing*. Hoboken, NJ: Wiley.

Løssing, A. S. W. (2008). *Danske kunstmuseer på nettet : en kortlægning og diskussion af en kunstmuseal formidlings- og udstillingspraksis.* PhD, Institut for Informations- og Medievidenskab, Aarhus Universitet, Århus.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge: Cambridge University Press.

Marchionini, G. (1995). *Information seeking in electronic environments*. Cambridge: Cambridge University Press.

Marchionini, G. (1999). Educating responsible citizens in the information society. *Educational Technology, 39,* 17-26.

Marchionini, G. (2006a). Exploratory Search: From Finding to Understanding. *Communications of the ACM, 49*(4), 41-46. doi: 10.1145/1121949.1121979

Marchionini, G. (2006b). Toward Human-Computer Information Retrieval. *Bulletin of the American Society for Information Science & Technology, 32*(5), 20-22. doi: 10.1002/bult.2006.1720320508

Marchionini, G. (2008). Human-information interaction research and development. *Library & Information Science Research, 30*(3), 165-174. doi: 10.1016/j.lisr.2008.07.001

Mat-Hassan, M., & Levene, M. (2005). Associating search and navigation behavior through log analysis. *Journal of the American Society for Information Science and Technology, 56*(9), 913-934. doi: 10.1002/asi.20185

McGrenere, J., & Ho, W. (2000). *Affordances: Clarifying and evolving a concept*. Paper presented at the Graphics Interface. doi: 10.1.1.153.8239

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis : an expanded sourcebook*. Thousand Oaks, CA: Sage.

Montgomery, A. L., Li, S., Srinivasan, K., & Liechty, J. C. (2004). Modeling Online Browsing and Path Analysis Using Clickstream Data. *Marketing Science, 23*(4), 579-595. doi: 10.1287/mksc.1040.0073

Morville, P. (2005). *Ambient findability*. Sebastopol, CA: O'Reilly.

Morville, P., & Rosenfeld, L. (2007). *Information architecture for the World Wide Web*. Sebastopol, CA: O'Reilly.

Nachmias, R., & Gilad, A. (2002). Needle in a hyperstack: Searching for information on the World Wide Web. *Journal of Research on Technology in Education, 34*(4), 475-486. doi: 10.1.1.138.8768

Neuman, W. L. (2006). *Social research methods : qualitative and quantitative approaches*. Boston, MA: Pearson/Allyn and Bacon.

Nicholas, D. (2009). Employing deep log analysis to evaluate the information-seeking behaviour of users of digital libraries. In G. Tsakonas & C. Papatheodorou (Eds.), *Evaluation of digital libraries : an insight into useful applications and methods* (pp. 121-146). Oxford: Chandos.

Nicholas, D., Huntington, P., Jamali, H. R., & Tenopir, C. (2006a). Finding information in (very large) digital libraries: a deep log approach to determining differences in use according to method of access. *The Journal of Academic Librarianship, 32*(2), 119-126. doi: 10.1016/j.acalib.2005.12.005

Nicholas, D., Huntington, P., Jamali, H. R., & Watkinson, A. (2006b). The information seeking behaviour of the users of digital scholarly journals. *Information Processing & Management, 42*(5), 1345-1365. doi: 10.1016/j.ipm.2006.02.001

Nicholas, D., Huntington, P., Williams, P., & Dobrowolski, T. (2006c). The digital information consumer. In A. Spink & C. Cole (Eds.), *New directions in human information behavior* (pp. 203-228). Dordrecht: Springer.

Nielsen, J. (2006). F-Shaped Pattern For Reading Web Content Retrieved 22/07/2013, 2013, from http://www.nngroup.com/articles/f-shaped-pattern-reading-web-content/

Nielsen, J., & Loranger, H. (2006). *Prioritizing Web usability*. Berkeley, CA: New Riders.

Norman, D. A. (1988). *The psychology of everyday things*. New York, NY: Basic Books.

Norman, D. A. (1999). Affordance, conventions, and design. *Interactions, 6*(3), 38-43. doi: 10.1145/301153.301168

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank Citation Ranking: Bringing Order to the Web: Stanford InfoLab.

Pariser, E. (2011). *The filter bubble : what the Internet is hiding from you*. New York: Penguin Press.

Peters, T. A. (1993). The history and development of transaction log analysis. *Library Hi Tech, 11*(2), 41-66. doi: 10.1108/eb047884

Petras, V. (2006). *Translating Dialects in Search: Mapping between Specialized Languages of Discourse and Documentary Languages*. PhD, University of California, Berkeley, Berkeley. Retrieved from http://people.ischool.berkeley.edu/~vivienp/diss/vpetras-dissertation2006-official.pdf

Pharo, N. (2002). *The SST method schema: A tool for analysing work task-based web information search processes*. PhD, University of Tampere, Tampere. Retrieved from http://acta.uta.fi/pdf/951-44-5355-7.pdf

Pharo, N. (2008). Småstegs søkeprosesser: hvordan "disjointed incrementalism" kan benyttes for å forstå informasjonssøkeatferd. *Dansk Biblioteksforskning, 4*(2), 33-40.

Pharo, N., & Järvelin, K. (2004). The SST method: a tool for analysing web information search processes. *Information Processing & Management, 40*(4), 633-654. doi: 10.1.1.135.2913

Pirolli, P. (2007). *Information foraging theory : adaptive interaction with information*. Oxford; New York: Oxford University Press.

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review, 106*(4), 643-675. doi: 10.1.1.31.5407

Plano Clark, V. L., & Badiee, M. (2010). Research questions in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Sage handbook of mixed methods in social & behavioral research* (pp. 275-304): Sage.

Pors, N. O. (1994). *Tilgængelighed og græsning : om bibliotekernes brugere, materialer og servicekvalitet*. Valby: Danmarks Biblioteksforenings Forlag.

Pors, N. O. (2005). Studerende, Google og biblioteker : om studerendes brug af biblioteker og informationsressourcer. København: Biblioteksstyrelsen.

Pors, N. O. (2011). Renewals and Interlibrary Loans in Libraries: An Analysis of Affordances and Changing User Behaviour. In S. Hiller, K. Justh, M. Kyrillidou & J. Self (Eds.), *Proceedings of the 2010 Library Assessment Conference. Building Effective, Sustainable, Practical Assessment, October 24 – 27 October 2010, Baltimore, Maryland.* Washington: Association of Research Libraries.

Prabha, C., Connaway, L. S., Olszewski, L., & Jenkins, L. R. (2007). What is enough? Satisficing information needs. *Journal of Documentation, 63*(1), 74-89. doi: 10.1108/00220410710723894

Rainie, L., & Jansen, B. J. (2009). Surveys as a complementary method for web log analysis. In B. J. Jansen, A. Spink & I. Taksa (Eds.), *Handbook of research on web log analysis* (pp. 39-64). Hersey: Information Science Research.

Rose, D. E., & Levinson, D. (2004). *Understanding user goals in web search*. Paper presented at the Proceedings of the 13th international conference on World Wide Web, New York, NY. doi: 10.1145/988672.988675

Rowlands, I., Nicholas, D., Williams, P., Huntington, P., Fieldhouse, M., Gunter, B., . . . Tenopir, C. (2008). *The Google generation: the information behaviour of the researcher of the future*. Paper presented at the Aslib Proceedings. doi: 10.1108/00012530810887953

Ruthven, I. (2011). Information Retrieval in Context. In M. Melucci & R. Baeza-Yates (Eds.), *Advanced Topics in Information Retrieval* (pp. 187-207). Berlin: Springer.

Sandstrom, P. E. (1994). An optimal foraging approach to information seeking and use. *Library Quarterly, 64*(4), 414-449. doi: 10.1086/602724

Sandstrom, P. E. (1999). Scholars as subsistence foragers. [Article]. *Bulletin of the American Society for Information Science and Technology, 25*(3), 17-20. doi: 10.1002/bult.116

Saracevic, T. (1996). *Modeling Interaction in Information Retrieval (IR): A Review and Proposal*. Paper presented at the Proceedings of the ASIS annual meeting.

Saracevic, T. (1997). *The stratified model of information retrieval interaction. Extension and applications*. Paper presented at the Proceedings of the 60th Annual Meeting of the American Society for Information Science.

Saracevic, T. (2009). Information Science *Encyclopedia of Library and Information Sciences, Third Edition* (pp. 2570-2585): Taylor & Francis. doi: 10.1081/e-elis3-120043704

Savolainen, R. (1995). Everyday life information seeking: approaching information seeking in the context of 'way of life'. *Library & Information Science Research, 17*(3), 259-294. doi: 10.1016/0740-8188(95)90048-9

Savolainen, R. (2007). Information Behavior and Information Practice: Reviewing the "Umbrella Concepts" of Information-Seeking Studies. *The Library Quarterly, 77*(2), 109-132.

Savolainen, R. (2008). *Everyday information practices : a social phenomenological perspective*. Lanham, MD: Scarecrow Press.

Savolainen, R. (2009). Everyday Life Information Seeking *Encyclopedia of Library and Information Sciences, Third Edition* (pp. 1780-1789): Taylor & Francis. doi: 10.1081/e-elis3-120043920

Savolainen, R., & Kari, J. (2004). Placing the Internet in information source horizons. A study of information seeking by Internet users in the context of self-development. *Library & Information Science Research, 26*(4), 415-433. doi: 10.1016/j.lisr.2004.04.004

Sherman, C., & Price, G. (2001). *The Invisible Web : uncovering information sources search engines can't see*. Medford, NJ: CyberAge Books.

Simon, H. (1971). Designing organizations for an information-rich world. In M. Greenberger (Ed.), *Computers, Communications and the Public Interest* (pp. 37-72). Baltimore, MD: Johns Hopkins University Press.

Skov, M. (2009). *The reinvented museum: exploring information seeking behaviour in a digital museum context.* PhD, Royal School of Library and Information Science, Copenhagen.

Skov, M., & Ingwersen, P. (2008). *Exploring information seeking behaviour in a digital museum context*. Paper presented at the Proceedings of the second international symposium on Information interaction in context. doi: 10.1145/1414694.1414719

Snickars, P. (2005). Arkiv, kulturarv och audiovisuella medier. In P. Aronsson & M. Hillström (Eds.), *Kulturarvens dynamik : det institutionaliserade kulturarvets förändringar* (pp. 209). Norrköping: Tema Kultur och samhälle, Campus Norrköping, Linköpings universitet.

Sonnenwald, D. H., & Wildemuth, B. M. (2001). Investigating Information Seeking Behavior Using the Concept of Information Horizons.

Sonnenwald, D. H., Wildemuth, B. S., & Harmon, G. L. (2001). A Research Method to Investigate Information Seeking using the Concept of Information Horizons: An Example from a Study of Lower Socio-economic Students' Information Seeking Behavior. *The New Review of Information Behavior Research, 2*, 65-86.

Spink, A., & Jansen, B. J. (2004). *Web search : public searching on the Web*. Boston, MA: Kluwer Academic Publishers.

Spool, J. M., Perfetti, C., & Brittan, D. (2004). *Designing for the scent of information*: User Interface Engineering.

Stebbins, R. A. (2007). *Serious leisure : a perspective for our time*. New Brunswick, NJ: Transaction Publishers.

Stebbins, R. A. (2009). Leisure and Its Relationship to Library and: Information Science: Bridging the Gap. *Library Trends, 57*(4), 618-631. doi: 10.1.1.225.935

Straub, K., & Weinschenk, S. (2003). Breadth vs. Depth: UI Design Newsletter – April, 2003 Retrieved 22/07/2013, 2013, from http://www.humanfactors.com/downloads/apr03.asp

Tague-Sutcliffe, J. (1992). An introduction to informetrics. *Information Processing & Management, 28*(1), 1-3. doi: 10.1016/0306-4573(92)90087-G

Taylor, R. S. (1968). Question-Negotiation and Information Seeking in Libraries. *College and Research Libraries, 29*(3), 178-194.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research : integrating quantitative and qualitative approaches in the social and behavioral sciences*. Los Angeles, CA: Sage.

Thatcher, J., Burks, M., Heilmann, C., Lawton Henry, S., Kirkpatrick, A., Lauke, P. H., . . . Waddell, C. D. (2006). *Web accessibility : web standards and regulatory compliance*. New York: FriendsofED.

The Apache Software Foundation. (2012). Apache HTTP Server Version 2.2: Access Log Retrieved 14/12, 2013, from http://httpd.apache.org/docs/2.2/logs.html#accesslog

The Dublin Core Metadata Initiative. (2012). Dublin Core Metadata Element Set, Version 1.1 Retrieved 24/10, 2013, from http://dublincore.org/documents/dces/

Thelwall, M. (2012). A History of Webometrics. [Article]. *Bulletin of the American Society for Information Science & Technology, 38*(6), 18-23.

Thelwall, M., Vaughan, L., & Björneborn, L. (2005). Webometrics. *Annual Review of Information Science and Technology, 39*(1), 81-135. doi: 10.1002/aris.1440390110

Theorizing digital cultural heritage : a critical discourse. (2007). Cambridge, Mass.: MIT Press.

Thurow, S., & Musica, N. (2009). *When search meets web usability*. Berkeley, CA: New Riders.

Turner, P., & Turner, S. (2009). Triangulation in practice. *Virtual Reality, 13*(3), 171-181. doi: 10.1007/s10055-009-0117-2

Tuten, T. L. (2010). Conducting online surveys. In S. Gosling & J. A. Johnson (Eds.), *Advanced methods for conducting online behavioral research* (pp. 179-192). Washington, DC: American Psychological Association.

Unesco. (1989). Draft medium-term plan, 1990-1995 : General Conference, Twenty-fifth session, Paris, 1989. Paris, France: Unesco.

Unesco. (2003). Charter on the Preservation of Digital Heritage Retrieved 13/3, 2014, from http://portal.unesco.org/en/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html

Unesco. (2008). Definition of the cultural heritage Retrieved Web Page, 2014, from http://www.unesco.org/new/en/culture/themes/illicit-trafficking-of-cultural-property/unesco-database-of-national-cultural-heritage-laws/frequently-asked-questions/definition-of-the-cultural-heritage/

W3C. Introduction to Understanding WCAG 2.0 Retrieved 19/10, 2013, from http://www.w3.org/TR/UNDERSTANDING-WCAG20/intro.html

W3C Network Working Group. (2005). RFC 3986 - Uniform Resource Identifier (URI): Generic Syntax, 2012, from http://www.rfc-base.org/rfc-3986.html

Walliman, N. (2011). *Research methods : the basics*. London; New York: Routledge.

Walter, A. (2008). *Building findable websites : web standards, SEO, and beyond*. Berkeley, CA: New Riders.

Weideman, M. (2009). *Website visibility : the theory and practice of improving rankings*. Oxford: Chandos Publishing.

Westergren, J. (2009, 17/6/2009). PageRank Retrieved 30/11, 2013, from http://www.seo-guide.se/pagerank

White, R. W., Marchionini, G., & Muresan, G. (2008). Evaluating exploratory search systems: Introduction to special topic issue of information processing and management. [Editorial]. *Information Processing & Management, 44*(2), 433-436. doi: 10.1016/j.ipm.2007.09.011

White, R. W., & Roth, R. A. (2009). *Exploratory search beyond the query-response paradigm*. San Rafael, CA: Morgan & Claypool. doi: 10.2200/S00174ED1V01Y200901ICR003

Wieland, J. L., Marslev, N., & Vestergaard, J. J. (n.d.). Kulturarven på nettet. København.

Wikimedia. (2012). GLAM/Case studies/British Museum Retrieved 20130512, from http://outreach.wikimedia.org/wiki/GLAM/Case_studies/British_Museum

Wikipedia. (2013). Findability Retrieved 19/10, 2013, from http://en.wikipedia.org/w/api.php?action=query&prop=revisions&titles=Findability&rvprop=timestamp; content&format=xml

Wilson, T. D. (1981). On user studies and information needs. *Journal of Documentation, 37*(1), 3. doi: 10.1108/00220410610714895

Wilson, T. D. (1997). Information behaviour: An interdisciplinary perspective. *Information Processing & Management, 33*(4), 551. doi: 10.1016/S0306-4573(97)00028-9

Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation, 55*(3), 249-270. doi: 10.1108/EUM0000000007145

Wilson, T. D. (2008). Review of: Savolainen, Reijo Everyday information practices: a social phenomenological perspective. Lanham, MD: Scarecrow Press, 2008. *Information Research, 14*(1).

Witten, I. H., Gori, M., & Numerico, T. (2006). *Web dragons : inside the myths of search engine technology*. Boston, MA: Morgan Kaufmann.

Wormell, I. (1985). *Subject access project - SAP : improved subject retrieval for monographic publications*. Lund.

Wyatt, L. (2010). Witty's Blog: The British Museum and Me Retrieved 20130512, from http://wittylama.com/2010/03/13/the-british-museum-and-me/

Xie, H. I. (2008). *Interactive information retrieval in digital environments*. Hershey: IGI Publishing.

Xie, H. I. (2009). Information Searching and Search Models *Encyclopedia of Library and Information Sciences, Third Edition* (pp. 2592-2604): Taylor & Francis. doi: 10.1081/e-elis3-120043745

Zhang, X., Anghelescu, H. G. B., & Yuan, X. (2005). Domain knowledge, search behaviour, and search effectiveness of engineering and science students: an exploratory study. *Information Research, 10*(2), 10-12.

Zipf, G. K. (1949). *Human behavior and the principle of least effort; an introduction to human ecology*. Cambridge, MA: Addison-Wesley Press.

List of abbreviations

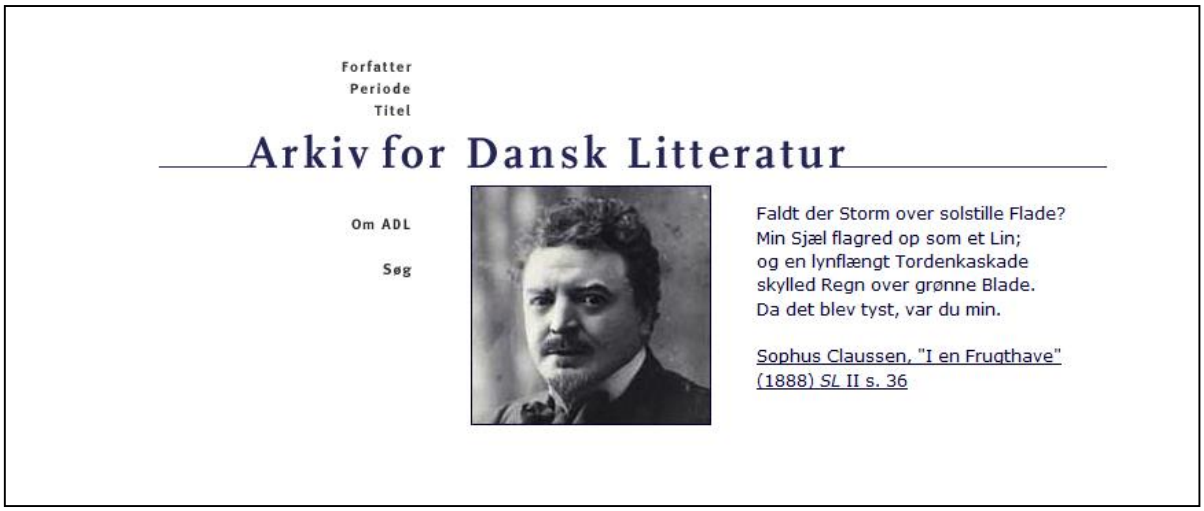
ADM	Alternative Document Models
ALM	Archives, Libraries and Museums
ATA	Access Target Area
CH	Cultural Heritage
HIB	Human Information Behaviour
I or I-level	Informational level or object
IA	Information Architecture
IIR	Interactive Information Retrieval
IR	Information Retrieval
IS	Information Seeking
IS&R	Information Seeking and Retrieval
N or N-level	Navigational level or object
O or O-level	Object level or Cultural Heritage object
SE	Search Engine
SEO	Search Engine Optimization
SERP	Search Engine Results Page
URI model	User-Resource Interaction model
URL	Uniform Resource Locator
WA	Web Analytics
WCAG	Web Content Accessibility Guidelines
WLS	Web Log Storming (software for log analysis)

Appendices

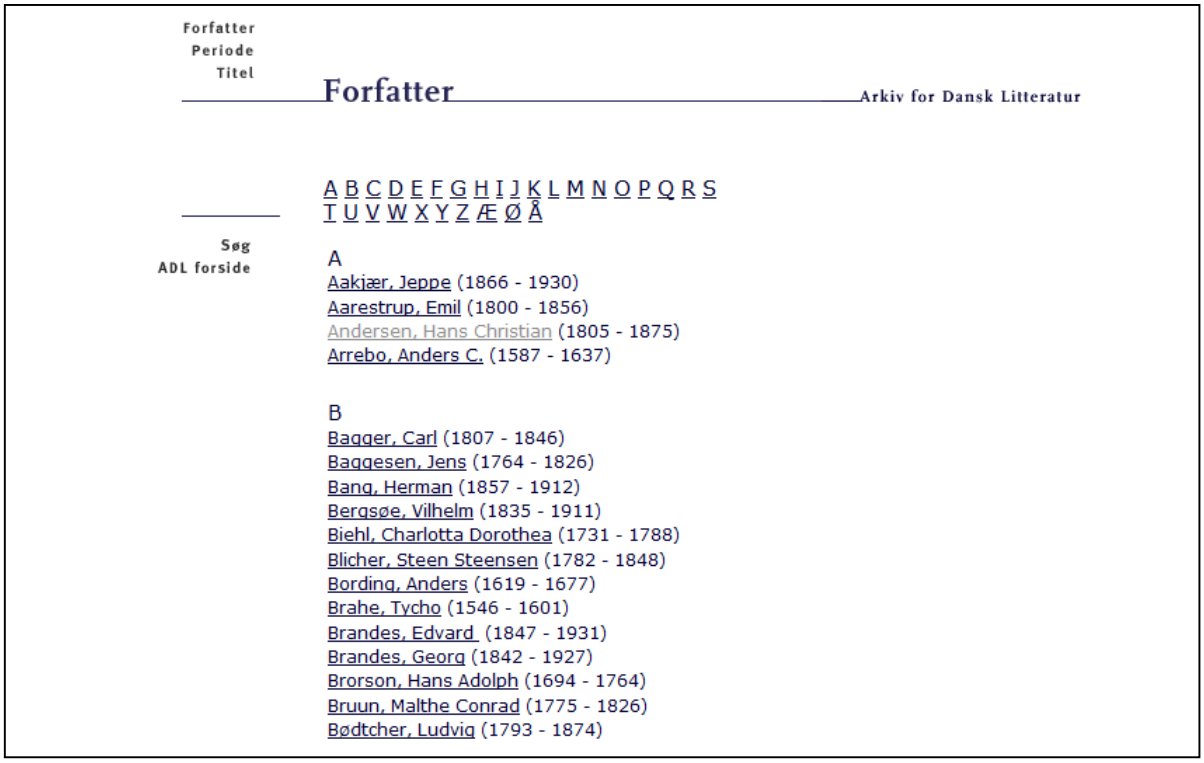
Appendix 1 Screen pictures of studied ADL-objects 204
Appendix 2 Screen pictures of studied KID-objects 210
Appendix 3 Screen pictures of studied Poma-objects 220
Appendix 4 Example of a session in ADL 224
Appendix 5 Screen shot of Web Log Storming 225
Appendix 6 Data about the sorting based on referrer in WLS 226
Appendix 7 Web survey questions 227
Appendix 8 Site structure analysis (including URL analysis) 233
Appendix 9 Referring search engines 237
Appendix 10 Object attributes 239
Appendix 11 Accessibility 242
Appendix 12 Internal navigation and internal search 245
Appendix 13 Reachability 248
Appendix 14 Web prestige 251
Appendix 15 Navigation strategies in logs 254
Appendix 16 Queries in referring search engines 255
Appendix 17 Referring sites (links group per site) 268
Appendix 18 Referring Wikipedia pages 271
Appendix 19 Countries of origin 273
Appendix 20 Distribution of session paths 274
Appendix 21 Distribution of session paths based on navigation strategy 276
Appendix 22 Navigation strategies, intentions and work contexts in survey 280
Appendix 23 Crosstabulations on survey data 281
Appendix 24 Comparison logs and survey 284

Appendix 1 Screen pictures of studied ADL-objects

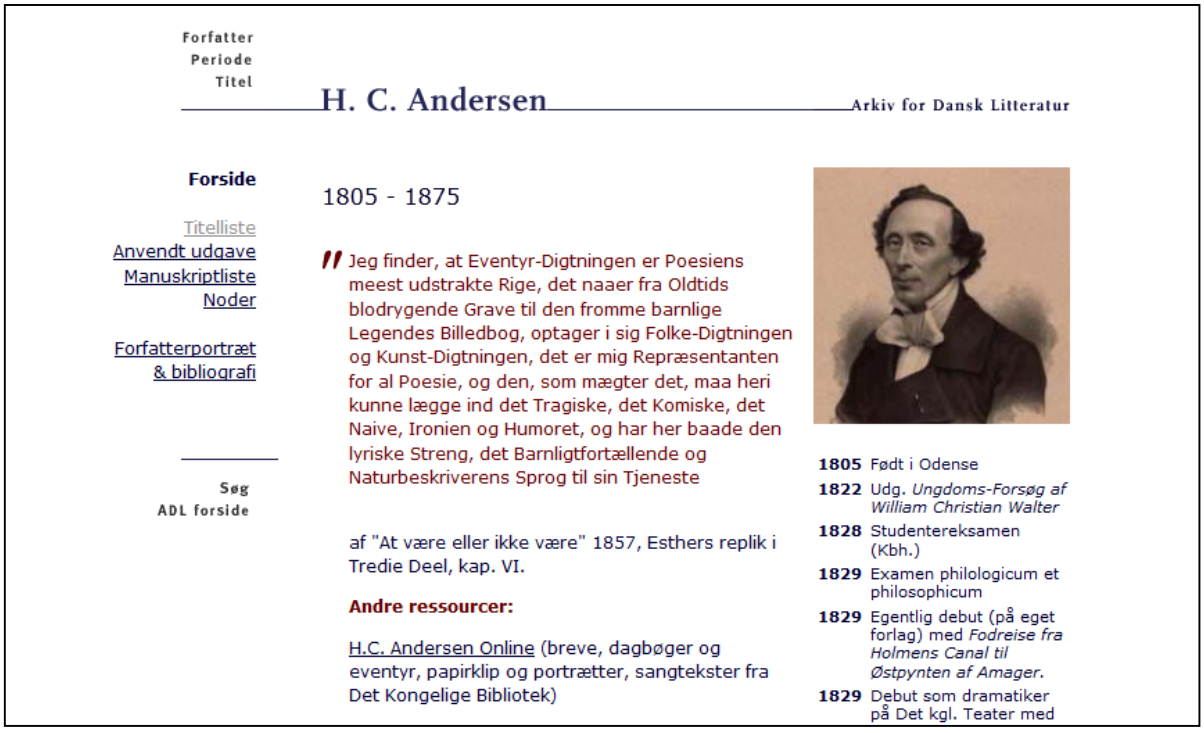
Examples of the studied objects, all objects are not included below.



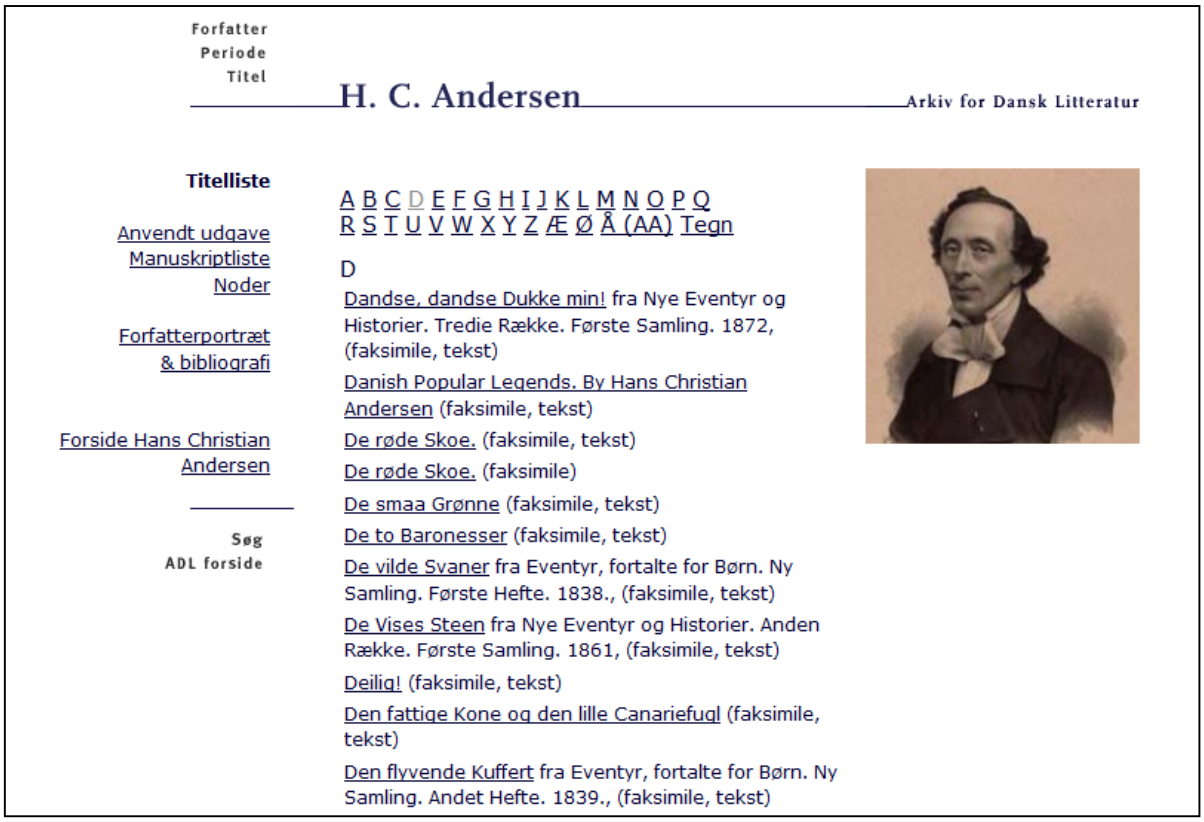
A-N1 (the picture and the quote changes regularly on the first page).



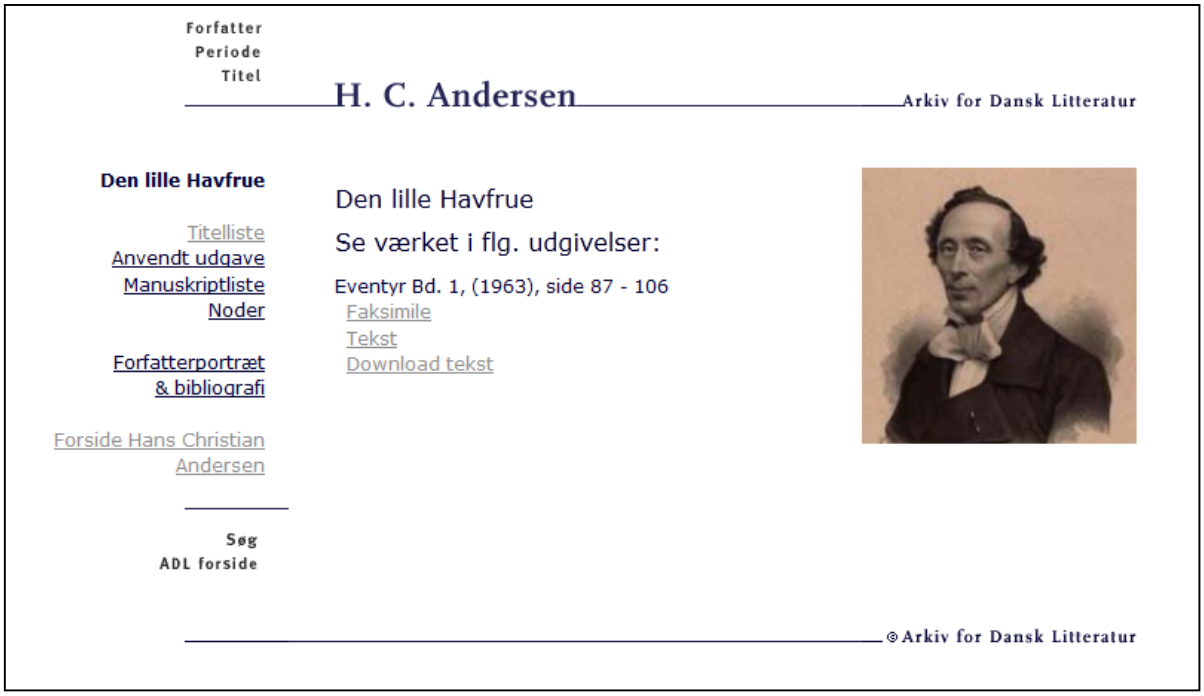
A-N2


Forfatter
Periode
Titel
H. C. Andersen
Arkiv for Dansk Litteratur
Forside
Titelliste
Anvendt udgave
Manuskriptliste
Noder
Forfatterportræt & bibliografi
Søg
ADL forside
1805 - 1875
Jeg finder, at Eventyr-Digtningen er Poesiens meest udstrakte Rige, det naaer fra Oldtids blodrygende Grave til den fromme barnlige Legendes Billedbog, optager i sig Folke-Digtningen og Kunst-Digtningen, det er mig Repræsentanten for al Poesie, og den, som mægter det, maa heri kunne lægge ind det Tragiske, det Komiske, det Naive, Ironien og Humoret, og har her baade den lyriske Streng, det Barnligtfortællende og Naturbeskriverens Sprog til sin Tjeneste
af "At være eller ikke være" 1857, Esthers replik i Tredie Deel, kap. VI.
Andre ressourcer:
H.C. Andersen Online (breve, dagbøger og eventyr, papirklip og portrætter, sangtekster fra Det Kongelige Bibliotek)
1805 Født i Odense
1822 Udg. Ungdoms-Forsøg af William Christian Walter
1828 Studentereksamen (Kbh.)
1829 Examen philologicum et philosophicum
1829 Egentlig debut (på eget forlag) med Fodreise fra Holmens Canal til Østpynten af Amager.
1829 Debut som dramatiker på Det kgl. Teater med

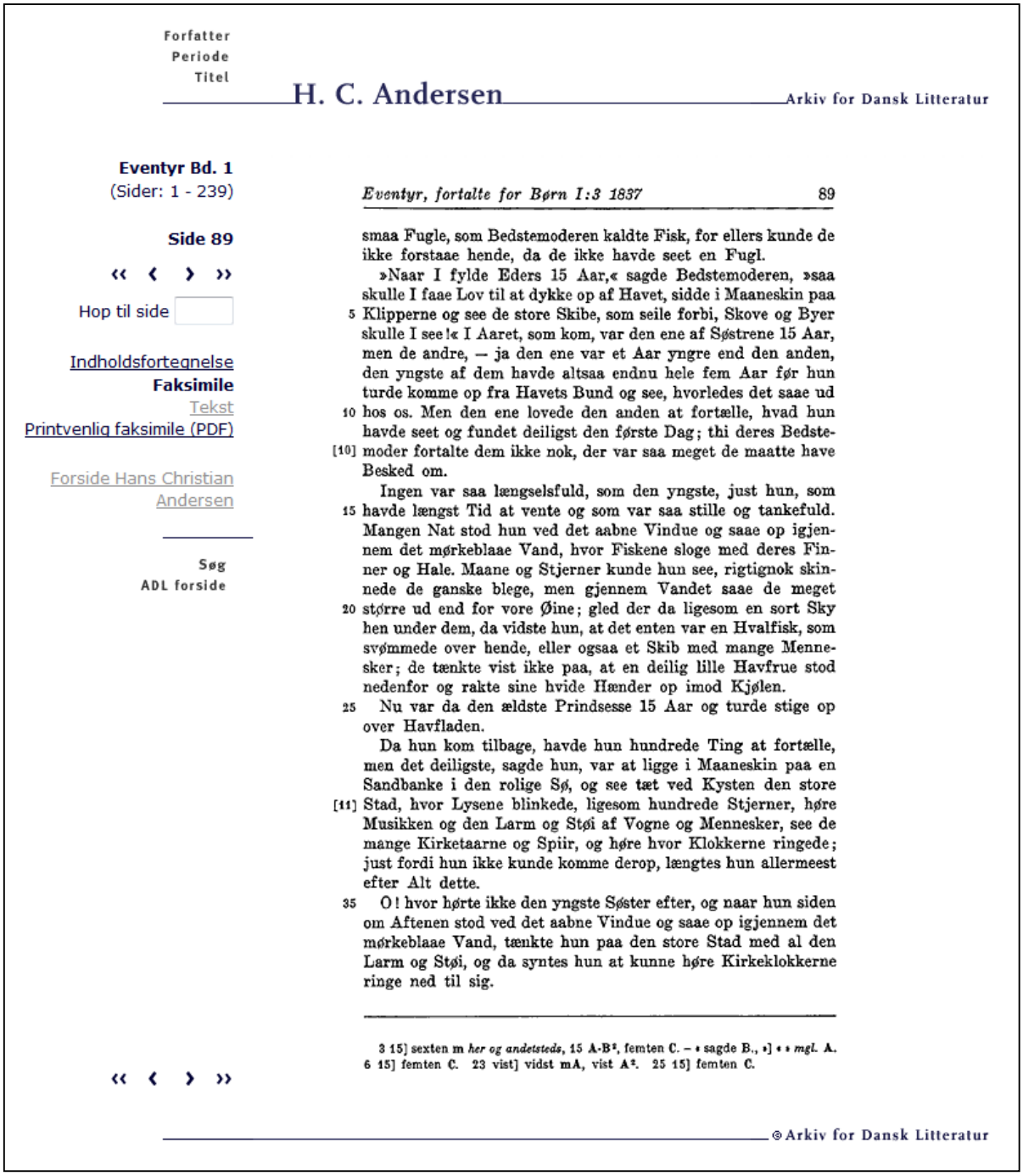
A-I1


Forfatter
Periode
Titel
H. C. Andersen
Arkiv for Dansk Litteratur
Titelliste
Anvendt udgave
Manuskriptliste
Noder
Forfatterportræt & bibliografi
Forside Hans Christian Andersen
Søg
ADL forside
A B C D E F G H I J K L M N O P Q
R S T U V W X Y Z Æ Ø Å (AA) Tegn
D
Dandse, dandse Dukke min! fra Nye Eventyr og Historier. Tredie Række. Første Samling. 1872, (faksimile, tekst)
Danish Popular Legends. By Hans Christian Andersen (faksimile, tekst)
De røde Skoe. (faksimile, tekst)
De røde Skoe. (faksimile)
De smaa Grønne (faksimile, tekst)
De to Baronesser (faksimile, tekst)
De vilde Svaner fra Eventyr, fortalte for Børn. Ny Samling. Første Hefte. 1838., (faksimile, tekst)
De Vises Steen fra Nye Eventyr og Historier. Anden Række. Første Samling. 1861, (faksimile, tekst)
Deilig! (faksimile, tekst)
Den fattige Kone og den lille Canariefugl (faksimile, tekst)
Den flyvende Kuffert fra Eventyr, fortalte for Børn. Ny Samling. Andet Hefte. 1839., (faksimile, tekst)

A-I2


Forfatter
Periode
Titel
H. C. Andersen
Arkiv for Dansk Litteratur
Den lille Havfrue
Titelliste
Anvendt udgave
Manuskriptliste
Noder
Forfatterportræt & bibliografi
Forside Hans Christian Andersen
Søg
ADL forside
Den lille Havfrue
Se værket i flg. udgivelser:
Eventyr Bd. 1, (1963), side 87 - 106
Faksimile
Tekst
Download tekst
© Arkiv for Dansk Litteratur

A-I3



Eventyr Bd. 1
(Sider: 1 - 239)

Side 89

« ‹ › »

Hop til side

Indholdsfortegnelse
Faksimile
Tekst
Printvenlig faksimile (PDF)

Forside Hans Christian Andersen

Søg
ADL forside

smaa Fugle, som Bedstemoderen kaldte Fisk, for ellers kunde de ikke forstaae hende, da de ikke havde seet en Fugl.

»Naar I fylde Eders 15 Aar,« sagde Bedstemoderen, »saa skulle I faae Lov til at dykke op af Havet, sidde i Maaneskin paa Klipperne og see de store Skibe, som seile forbi, Skove og Byer skulle I see!« I Aaret, som kom, var den ene af Søstrene 15 Aar, men de andre, — ja den ene var et Aar yngre end den anden, den yngste af dem havde altsaa endnu hele fem Aar før hun turde komme op fra Havets Bund og see, hvorledes det saae ud hos os. Men den ene lovede den anden at fortælle, hvad hun havde seet og fundet deiligst den første Dag; thi deres Bedste- [10] moder fortalte dem ikke nok, der var saa meget de maatte have Besked om.

Ingen var saa længselsfuld, som den yngste, just hun, som havde længst Tid at vente og som var saa stille og tankefuld. Mangen Nat stod hun ved det aabne Vindue og saae op igjennem det mørkeblaae Vand, hvor Fiskene sloge med deres Finner og Hale. Maane og Stjerner kunde hun see, rigtignok skinnede de ganske blege, men gjennem Vandet saae de meget større ud end for vore Øine; gled der da ligesom en sort Sky hen under dem, da vidste hun, at det enten var en Hvalfisk, som svømmede over hende, eller ogsaa et Skib med mange Mennesker; de tænkte vist ikke paa, at en deilig lille Havfrue stod nedenfor og rakte sine hvide Hænder op imod Kjølen.

Nu var da den ældste Prindsesse 15 Aar og turde stige op over Havfladen.

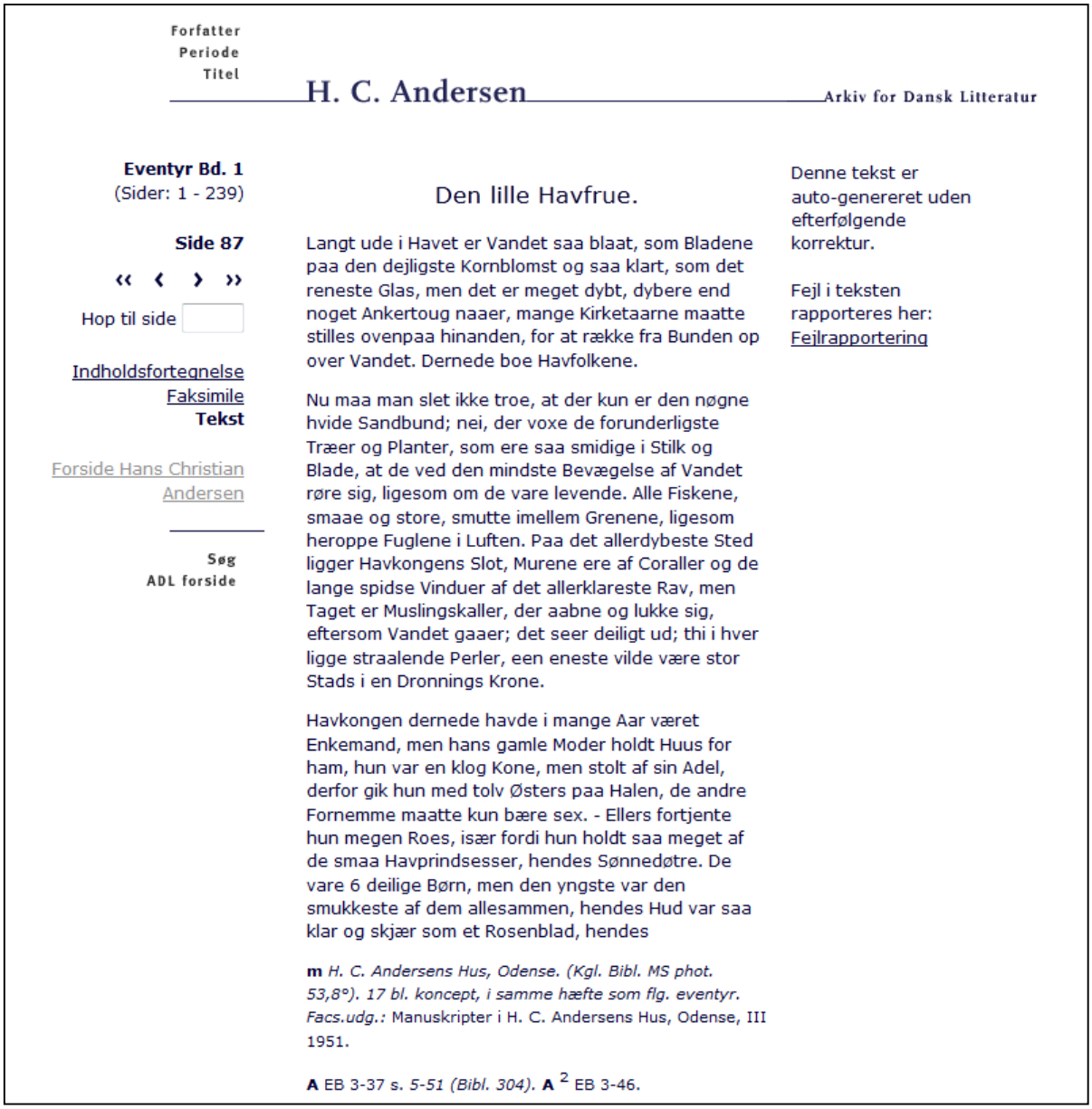
Da hun kom tilbage, havde hun hundrede Ting at fortælle, men det deiligste, sagde hun, var at ligge i Maaneskin paa en Sandbanke i den rolige Sø, og see tæt ved Kysten den store [11] Stad, hvor Lysene blinkede, ligesom hundrede Stjerner, høre Musikken og den Larm og Støi af Vogne og Mennesker, see de mange Kirketaarne og Spiir, og høre hvor Klokkerne ringede; just fordi hun ikke kunde komme derop, længtes hun allermeest efter Alt dette.

O! hvor hørte ikke den yngste Søster efter, og naar hun siden om Aftenen stod ved det aabne Vindue og saae op igjennem det mørkeblaae Vand, tænkte hun paa den store Stad med al den Larm og Støi, og da syntes hun at kunne høre Kirkeklokkerne ringe ned til sig.

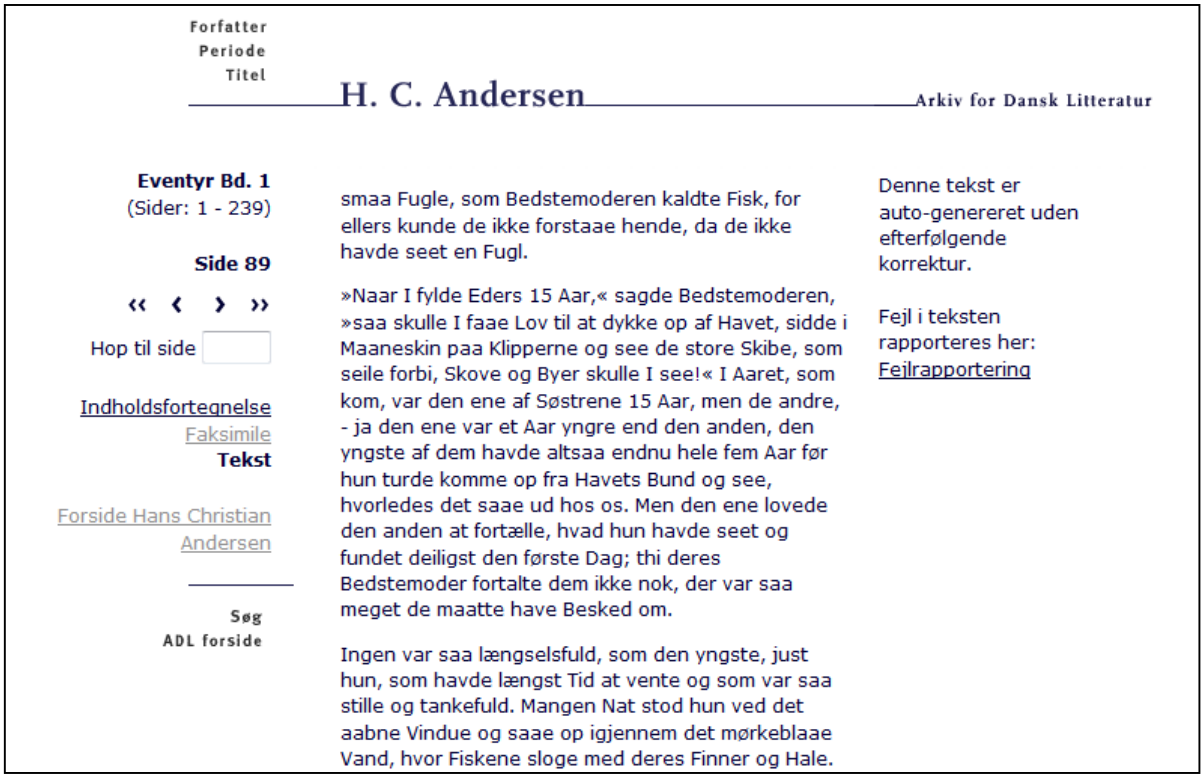
3 15] sexten m *her og andetsteds*, 15 A-B², femten C. – ∘ sagde B., ∘] ∘ ∘ *mgl.* A. 6 15] femten C. 23 vist] vidst mA, vist A². 25 15] femten C.

« ‹ › »

A-O2


Forfatter
Periode
Titel
H. C. Andersen
Arkiv for Dansk Litteratur
Eventyr Bd. 1
(Sider: 1 - 239)
Side 87
Hop til side
Indholdsfortegnelse
Faksimile
Tekst
Forside Hans Christian Andersen
Søg
ADL forside
Den lille Havfrue.
Langt ude i Havet er Vandet saa blaat, som Bladene paa den dejligste Kornblomst og saa klart, som det reneste Glas, men det er meget dybt, dybere end noget Ankertoug naaer, mange Kirketaarne maatte stilles ovenpaa hinanden, for at række fra Bunden op over Vandet. Dernede boe Havfolkene.
Nu maa man slet ikke troe, at der kun er den nøgne hvide Sandbund; nei, der voxe de forunderligste Træer og Planter, som ere saa smidige i Stilk og Blade, at de ved den mindste Bevægelse af Vandet røre sig, ligesom om de vare levende. Alle Fiskene, smaae og store, smutte imellem Grenene, ligesom heroppe Fuglene i Luften. Paa det allerdybeste Sted ligger Havkongens Slot, Murene ere af Coraller og de lange spidse Vinduer af det allerklareste Rav, men Taget er Muslingskaller, der aabne og lukke sig, eftersom Vandet gaaer; det seer deiligt ud; thi i hver ligge straalende Perler, een eneste vilde være stor Stads i en Dronnings Krone.
Havkongen dernede havde i mange Aar været Enkemand, men hans gamle Moder holdt Huus for ham, hun var en klog Kone, men stolt af sin Adel, derfor gik hun med tolv Østers paa Halen, de andre Fornemme maatte kun bære sex. - Ellers fortjente hun megen Roes, især fordi hun holdt saa meget af de smaa Havprindsesser, hendes Sønnedøtre. De vare 6 deilige Børn, men den yngste var den smukkeste af dem allesammen, hendes Hud var saa klar og skjær som et Rosenblad, hendes
m H. C. Andersens Hus, Odense. (Kgl. Bibl. MS phot. 53,8°). 17 bl. koncept, i samme hæfte som flg. eventyr. Facs.udg.: Manuskripter i H. C. Andersens Hus, Odense, III 1951.
A EB 3-37 s. 5-51 (Bibl. 304). A 2 EB 3-46.
Denne tekst er auto-genereret uden efterfølgende korrektur.
Fejl i teksten rapporteres her:
Fejlrapportering

A-O3



Forfatter
Periode
Titel

H. C. Andersen

Arkiv for Dansk Litteratur

Eventyr Bd. 1
(Sider: 1 - 239)

Side 89

‹‹ ‹ › ››

Hop til side

Indholdsfortegnelse
Faksimile
Tekst

Forside Hans Christian Andersen

Søg
ADL forside

smaa Fugle, som Bedstemoderen kaldte Fisk, for ellers kunde de ikke forstaae hende, da de ikke havde seet en Fugl.

»Naar I fylde Eders 15 Aar,« sagde Bedstemoderen, »saa skulle I faae Lov til at dykke op af Havet, sidde i Maaneskin paa Klipperne og see de store Skibe, som seile forbi, Skove og Byer skulle I see!« I Aaret, som kom, var den ene af Søstrene 15 Aar, men de andre, - ja den ene var et Aar yngre end den anden, den yngste af dem havde altsaa endnu hele fem Aar før hun turde komme op fra Havets Bund og see, hvorledes det saae ud hos os. Men den ene lovede den anden at fortælle, hvad hun havde seet og fundet deiligst den første Dag; thi deres Bedstemoder fortalte dem ikke nok, der var saa meget de maatte have Besked om.

Ingen var saa længselsfuld, som den yngste, just hun, som havde længst Tid at vente og som var saa stille og tankefuld. Mangen Nat stod hun ved det aabne Vindue og saae op igjennem det mørkeblaae Vand, hvor Fiskene sloge med deres Finner og Hale.

Denne tekst er auto-genereret uden efterfølgende korrektur.

Fejl i teksten rapporteres her:
Fejlrapportering

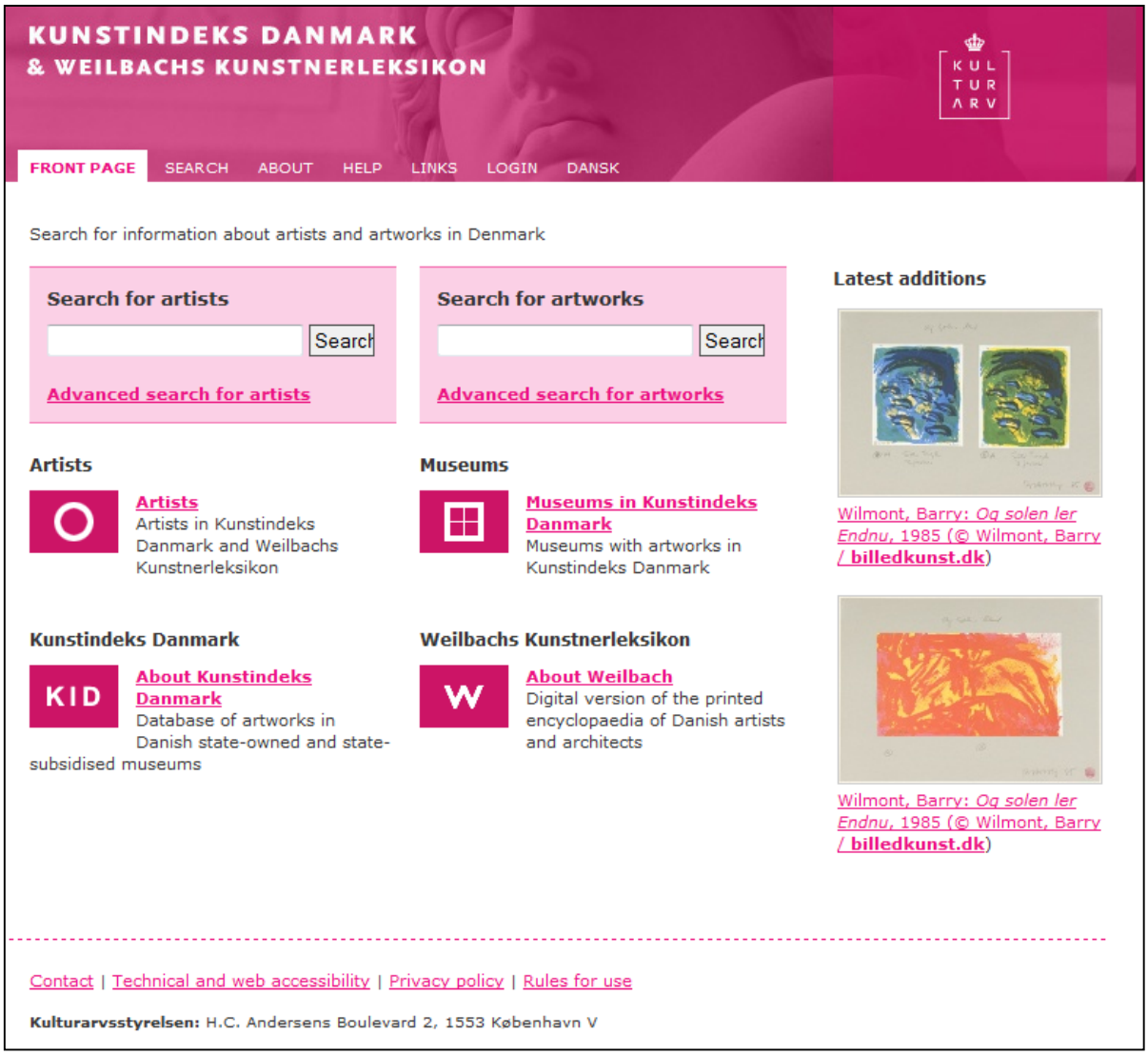
A-O4

```
[ Arkiv for Dansk Litteratur - ascii-version af værk. NB: teksten er autogenereret uden korrektur !! ]


Den lille Havfrue.

Langt ude i Havet er Vandet saa blaat, som Bladene paa den dejligste Kornblomst og saa klart, som det rer

Nu maa man slet ikke troe, at der kun er den nøgne hvide Sandbund; nei, der voxe de forunderligste Træer

Havkongen dernede havde i mange Aar været Enkemand, men hans gamle Moder holdt Huus for ham, hun var en k

mH. C. Andersens Hus, Odense. (Kgl. Bibl. MS phot. 53,8°). 17 bl. koncept,i samme hæfte som flg. eventyr.

AEB 3-37 s.5-51 (Bibl. 304).A2EB 3-46.

BEP-50 s.166-95 m. 4 ill., deraf 2 på indsatte blade.B2EP-54.

CSS 19-55s. 124-48.

DEHP 1-62s. 100-34 m. 4 ill.

23 6] sexC.


Øine saa blaa, som den dybeste Sø, men ligesom alle de andre havde hun ingen Fødder, Kroppen endte i en F

Hele den lange Dag kunde de lege nede i Slottet, i de store Sale, hvor levende Blomster voxte ud af Vægge

Udenfor Slottet var en stor Have med ildrøde og mørkeblaae Træer, Frugterne straalede som Guld, og Blomst
```

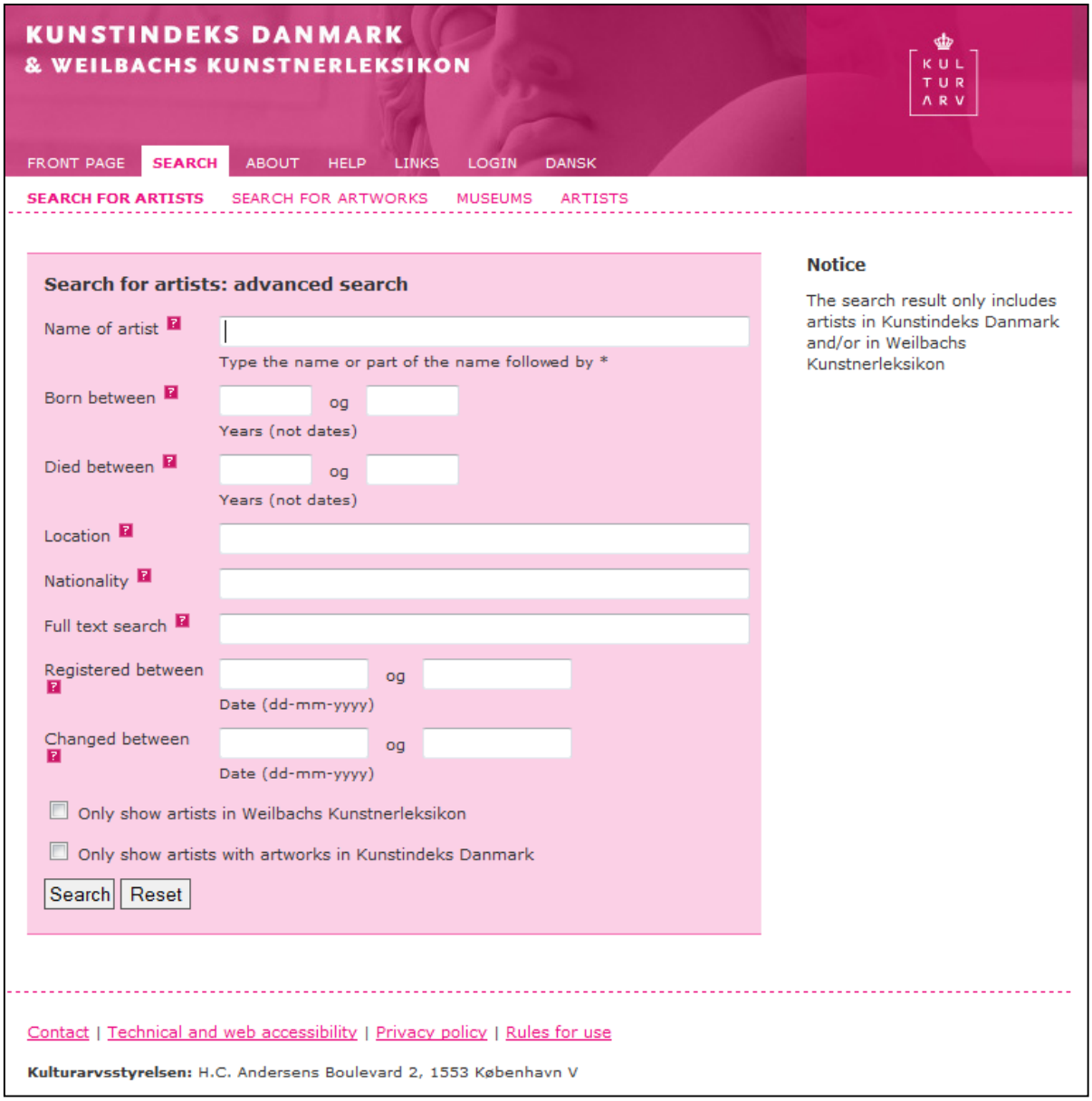
A-O5

Appendix 2 Screen pictures of studied KID-objects

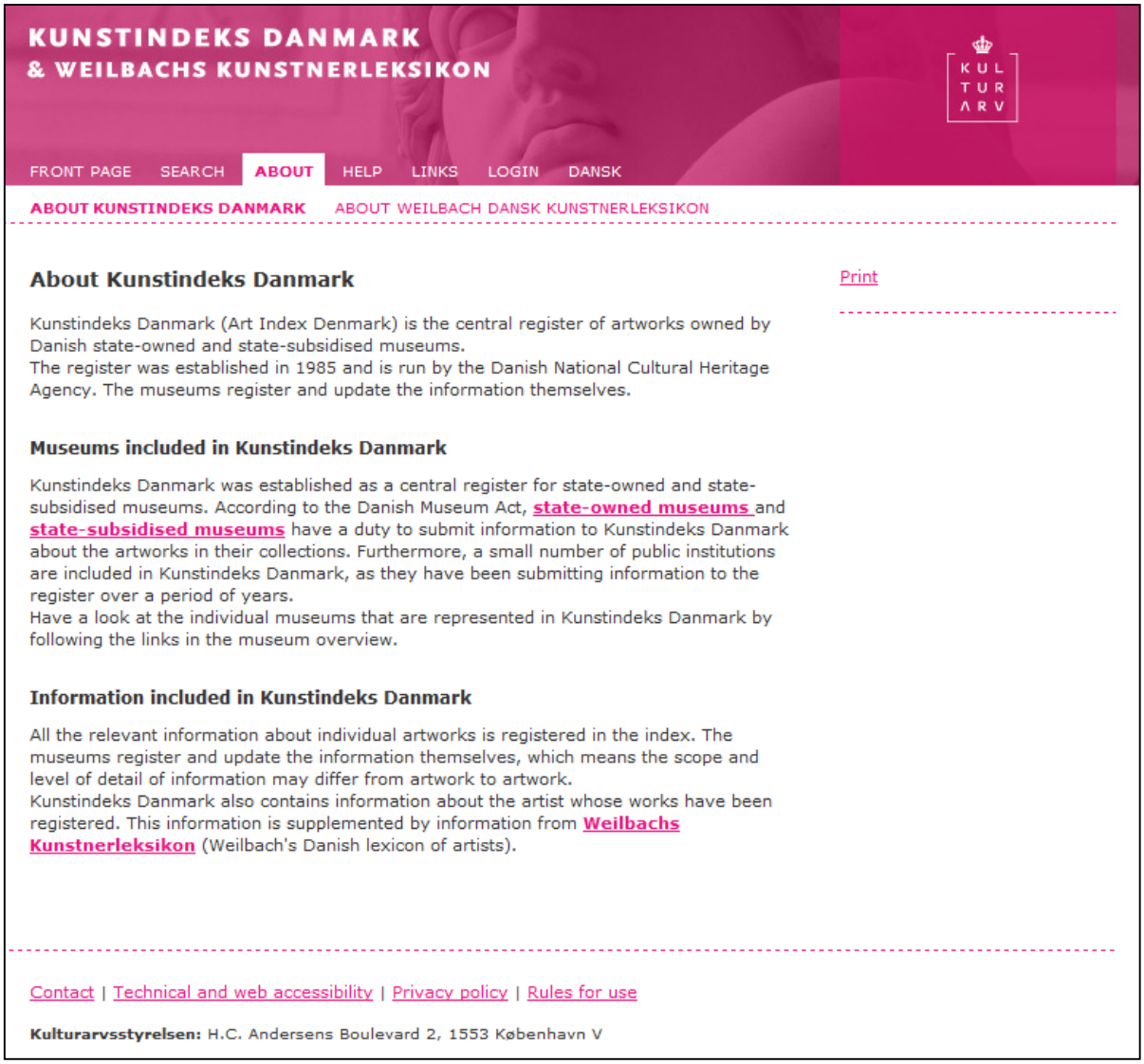
Examples of the studied objects, all objects are not included below.



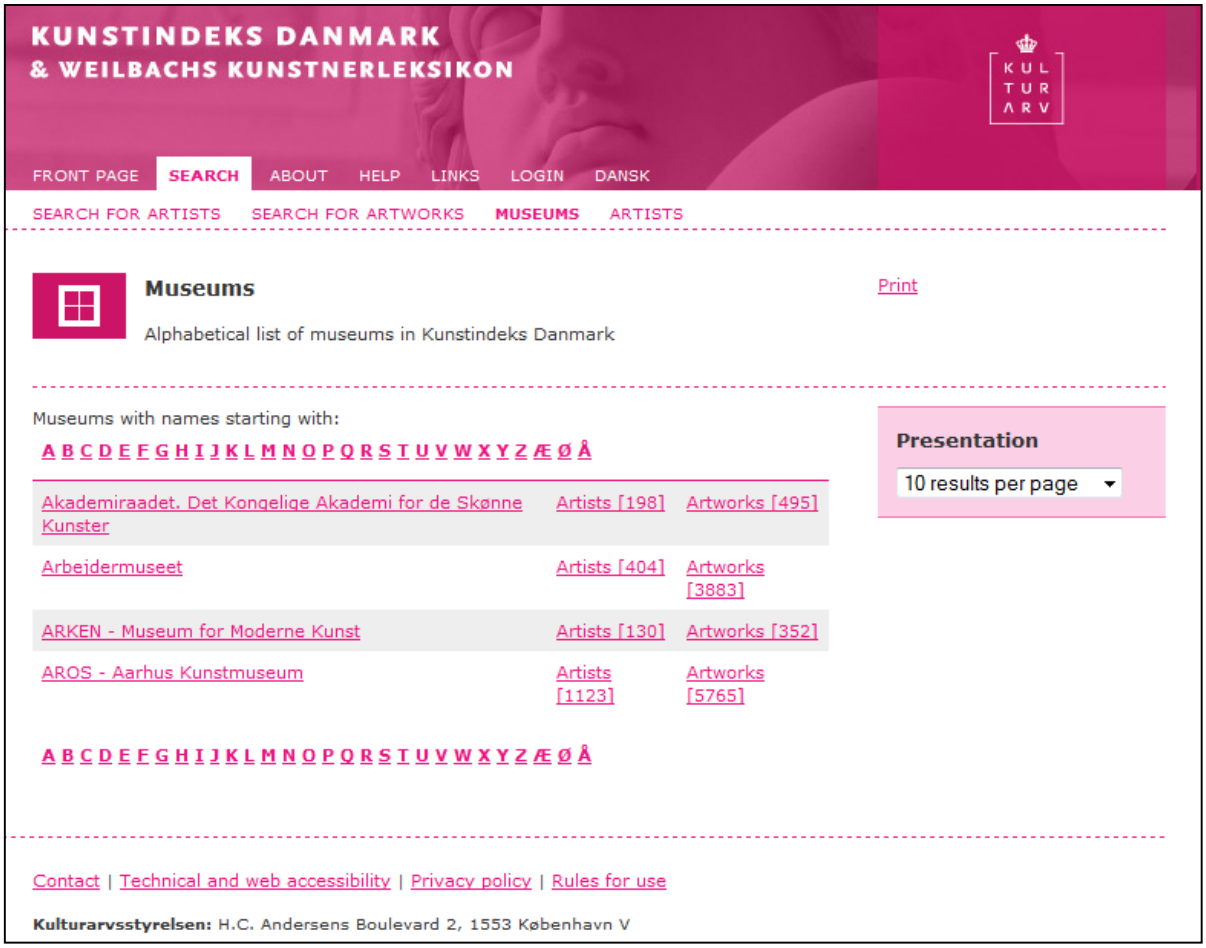
K-N1


KUNSTINDEKS DANMARK
& WEILBACHS KUNSTNERLEKSIKON
KUL TUR ARV
FRONT PAGE
SEARCH
ABOUT
HELP
LINKS
LOGIN
DANSK
SEARCH FOR ARTISTS
SEARCH FOR ARTWORKS
MUSEUMS
ARTISTS
Search for artists: advanced search
Name of artist
Type the name or part of the name followed by *
Born between
og
Years (not dates)
Died between
og
Years (not dates)
Location
Nationality
Full text search
Registered between
og
Date (dd-mm-yyyy)
Changed between
og
Date (dd-mm-yyyy)
Only show artists in Weilbachs Kunstnerleksikon
Only show artists with artworks in Kunstindeks Danmark
Search
Reset
Notice
The search result only includes artists in Kunstindeks Danmark and/or in Weilbachs Kunstnerleksikon
Contact | Technical and web accessibility | Privacy policy | Rules for use
Kulturarvsstyrelsen: H.C. Andersens Boulevard 2, 1553 København V

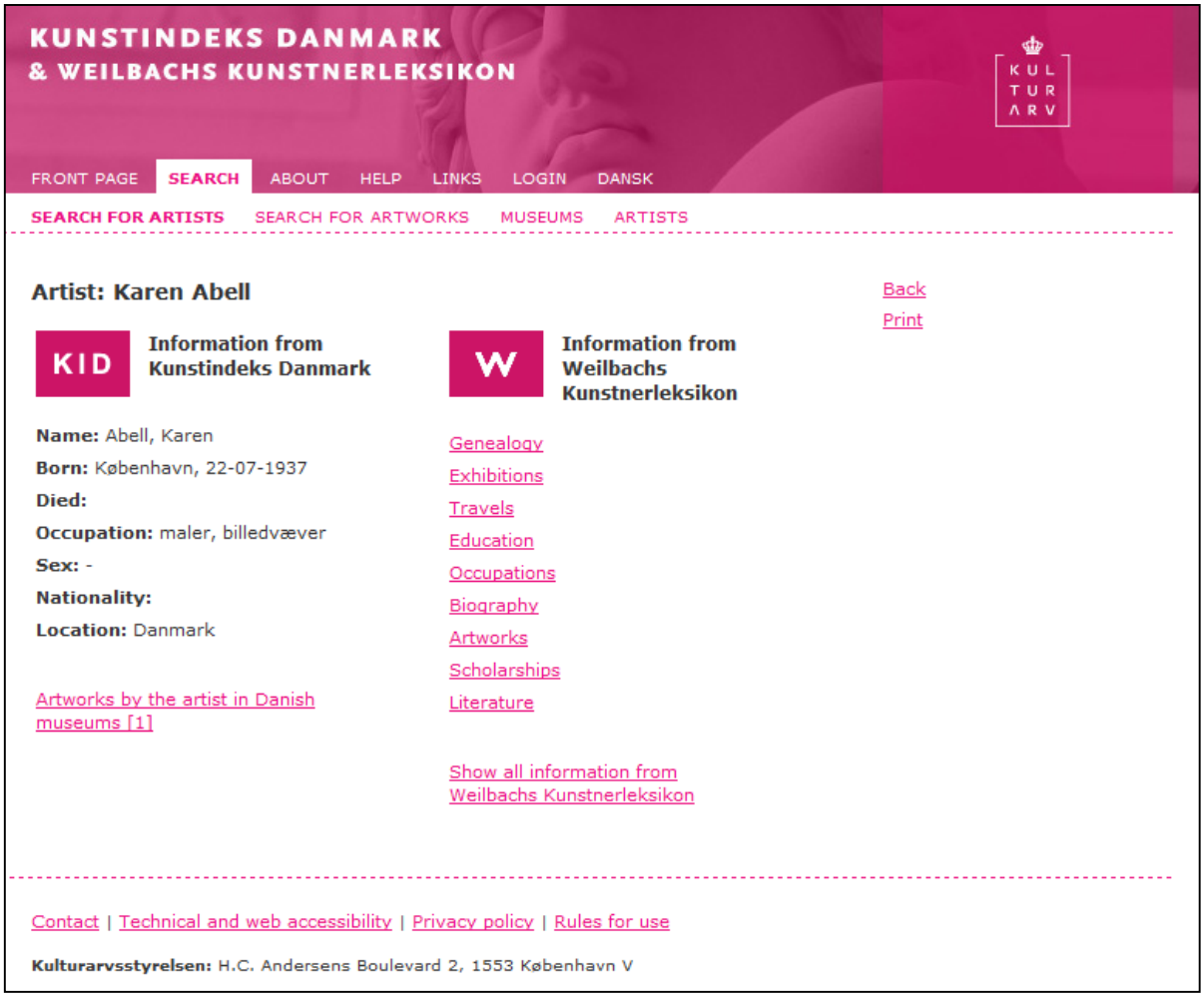
K-N2


KUNSTINDEKS DANMARK
& WEILBACHS KUNSTNERLEKSIKON
KUL TUR ARV
FRONT PAGE SEARCH ABOUT HELP LINKS LOGIN DANSK
ABOUT KUNSTINDEKS DANMARK ABOUT WEILBACH DANSK KUNSTNERLEKSIKON
About Kunstindeks Danmark
Print
Kunstindeks Danmark (Art Index Denmark) is the central register of artworks owned by Danish state-owned and state-subsidised museums.
The register was established in 1985 and is run by the Danish National Cultural Heritage Agency. The museums register and update the information themselves.
Museums included in Kunstindeks Danmark
Kunstindeks Danmark was established as a central register for state-owned and state-subsidised museums. According to the Danish Museum Act, state-owned museums and state-subsidised museums have a duty to submit information to Kunstindeks Danmark about the artworks in their collections. Furthermore, a small number of public institutions are included in Kunstindeks Danmark, as they have been submitting information to the register over a period of years.
Have a look at the individual museums that are represented in Kunstindeks Danmark by following the links in the museum overview.
Information included in Kunstindeks Danmark
All the relevant information about individual artworks is registered in the index. The museums register and update the information themselves, which means the scope and level of detail of information may differ from artwork to artwork.
Kunstindeks Danmark also contains information about the artist whose works have been registered. This information is supplemented by information from Weilbachs Kunstnerleksikon (Weilbach's Danish lexicon of artists).
Contact | Technical and web accessibility | Privacy policy | Rules for use
Kulturarvsstyrelsen: H.C. Andersens Boulevard 2, 1553 København V

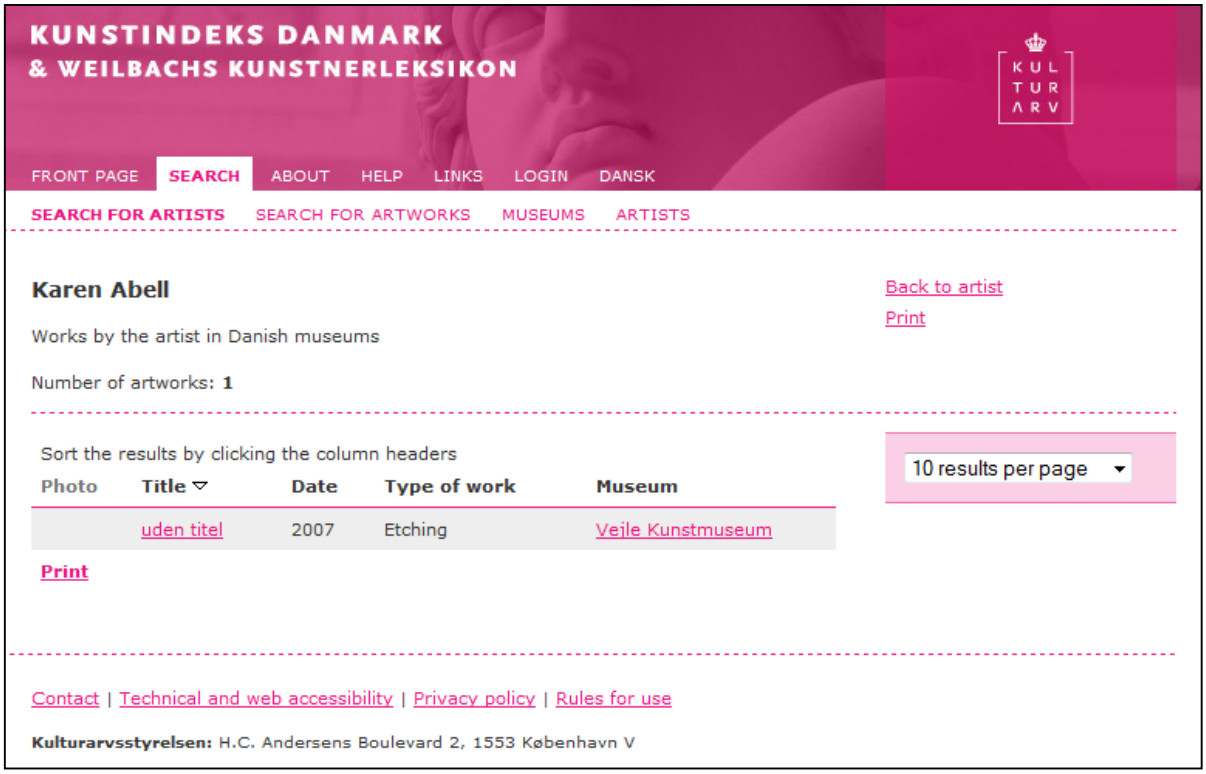
K-N3


KUNSTINDEKS DANMARK
& WEILBACHS KUNSTNERLEKSIKON
KUL TUR ARV
FRONT PAGE
SEARCH
ABOUT
HELP
LINKS
LOGIN
DANSK
SEARCH FOR ARTISTS
SEARCH FOR ARTWORKS
MUSEUMS
ARTISTS
Museums
Alphabetical list of museums in Kunstindeks Danmark
Print
Museums with names starting with:
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z Æ Ø Å
Presentation
10 results per page
Akademiraadet. Det Kongelige Akademi for de Skønne Kunster
Artists [198]
Artworks [495]
Arbejdermuseet
Artists [404]
Artworks [3883]
ARKEN - Museum for Moderne Kunst
Artists [130]
Artworks [352]
AROS - Aarhus Kunstmuseum
Artists [1123]
Artworks [5765]
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z Æ Ø Å
Contact | Technical and web accessibility | Privacy policy | Rules for use
Kulturarvsstyrelsen: H.C. Andersens Boulevard 2, 1553 København V

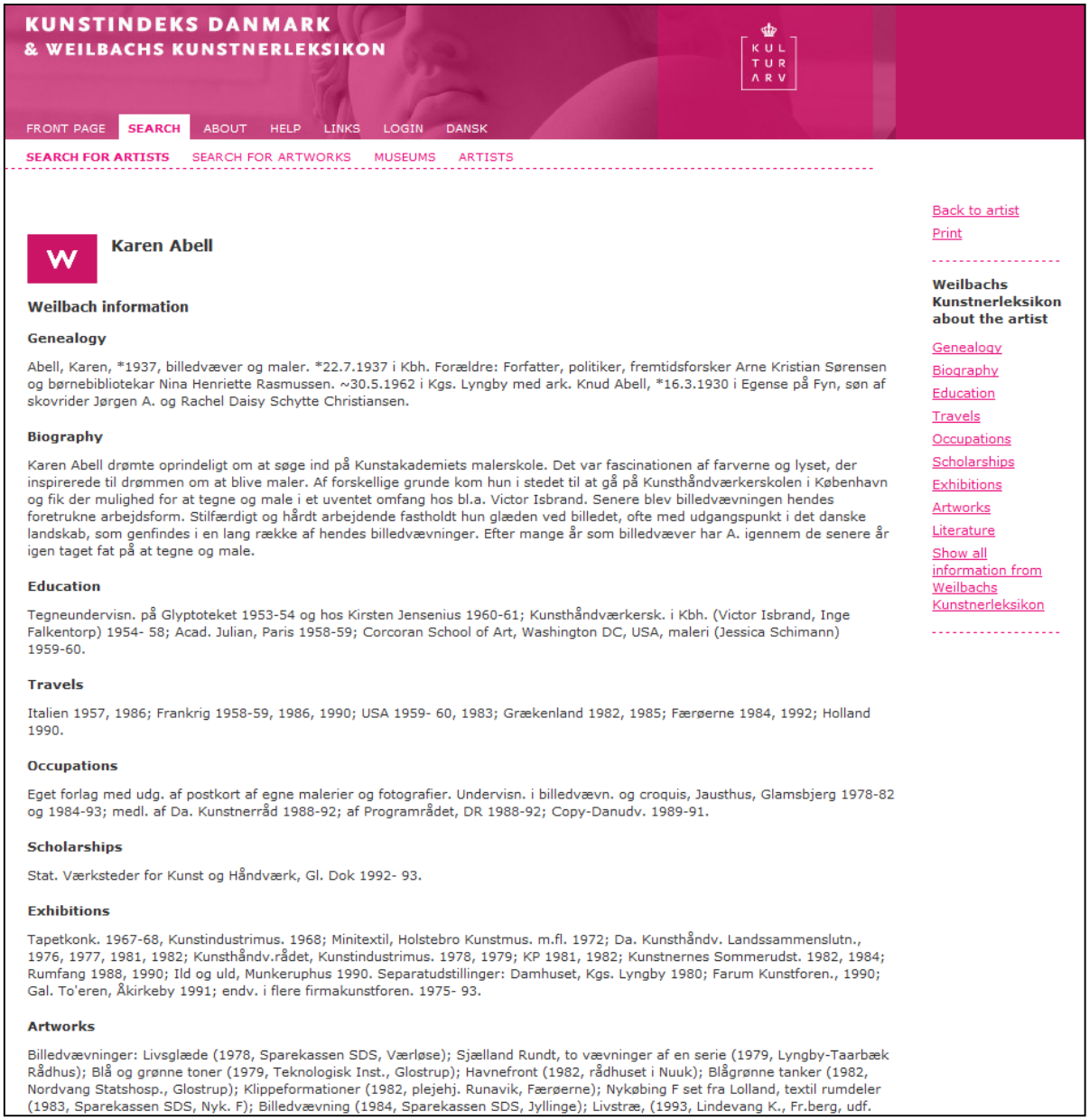
K-N4


KUNSTINDEKS DANMARK
& WEILBACHS KUNSTNERLEKSIKON
KUL TUR ARV
FRONT PAGE
SEARCH
ABOUT
HELP
LINKS
LOGIN
DANSK
SEARCH FOR ARTISTS
SEARCH FOR ARTWORKS
MUSEUMS
ARTISTS
Artist: Karen Abell
Back
Print
KID
Information from Kunstindeks Danmark
W
Information from Weilbachs Kunstnerleksikon
Name: Abell, Karen
Born: København, 22-07-1937
Died:
Occupation: maler, billedvæver
Sex: -
Nationality:
Location: Danmark
Artworks by the artist in Danish museums [1]
Genealogy
Exhibitions
Travels
Education
Occupations
Biography
Artworks
Scholarships
Literature
Show all information from Weilbachs Kunstnerleksikon
Contact | Technical and web accessibility | Privacy policy | Rules for use
Kulturarvsstyrelsen: H.C. Andersens Boulevard 2, 1553 København V

K-I1


KUNSTINDEKS DANMARK
& WEILBACHS KUNSTNERLEKSIKON
KUL TUR ARV
FRONT PAGE
SEARCH
ABOUT
HELP
LINKS
LOGIN
DANSK
SEARCH FOR ARTISTS
SEARCH FOR ARTWORKS
MUSEUMS
ARTISTS
Karen Abell
Back to artist
Print
Works by the artist in Danish museums
Number of artworks: 1
Sort the results by clicking the column headers
10 results per page
Photo
Title
Date
Type of work
Museum
uden titel
2007
Etching
Vejle Kunstmuseum
Print
Contact | Technical and web accessibility | Privacy policy | Rules for use
Kulturarvsstyrelsen: H.C. Andersens Boulevard 2, 1553 København V

K-I2



KUNSTINDEKS DANMARK
& WEILBACHS KUNSTNERLEKSIKON

FRONT PAGE | SEARCH | ABOUT | HELP | LINKS | LOGIN | DANSK

SEARCH FOR ARTISTS | SEARCH FOR ARTWORKS | MUSEUMS | ARTISTS

Back to artist
Print

Weilbachs Kunstnerleksikon about the artist

Genealogy
Biography
Education
Travels
Occupations
Scholarships
Exhibitions
Artworks
Literature
Show all information from Weilbachs Kunstnerleksikon

Karen Abell

Weilbach information

Genealogy

Abell, Karen, *1937, billedvæver og maler. *22.7.1937 i Kbh. Forældre: Forfatter, politiker, fremtidsforsker Arne Kristian Sørensen og børnebibliotekar Nina Henriette Rasmussen. ~30.5.1962 i Kgs. Lyngby med ark. Knud Abell, *16.3.1930 i Egense på Fyn, søn af skovrider Jørgen A. og Rachel Daisy Schytte Christiansen.

Biography

Karen Abell drømte oprindeligt om at søge ind på Kunstakademiets malerskole. Det var fascinationen af farverne og lyset, der inspirerede til drømmen om at blive maler. Af forskellige grunde kom hun i stedet til at gå på Kunsthåndværkerskolen i København og fik der mulighed for at tegne og male i et uventet omfang hos bl.a. Victor Isbrand. Senere blev billedvævningen hendes foretrukne arbejdsform. Stilfærdigt og hårdt arbejdende fastholdt hun glæden ved billedet, ofte med udgangspunkt i det danske landskab, som genfindes i en lang række af hendes billedvævninger. Efter mange år som billedvæver har A. igennem de senere år igen taget fat på at tegne og male.

Education

Tegneundervisn. på Glyptoteket 1953-54 og hos Kirsten Jensenius 1960-61; Kunsthåndværkersk. i Kbh. (Victor Isbrand, Inge Falkentorp) 1954- 58; Acad. Julian, Paris 1958-59; Corcoran School of Art, Washington DC, USA, maleri (Jessica Schimann) 1959-60.

Travels

Italien 1957, 1986; Frankrig 1958-59, 1986, 1990; USA 1959- 60, 1983; Grækenland 1982, 1985; Færøerne 1984, 1992; Holland 1990.

Occupations

Eget forlag med udg. af postkort af egne malerier og fotografier. Undervisn. i billedvævn. og croquis, Jausthus, Glamsbjerg 1978-82 og 1984-93; medl. af Da. Kunstnerråd 1988-92; af Programrådet, DR 1988-92; Copy-Danudv. 1989-91.

Scholarships

Stat. Værksteder for Kunst og Håndværk, Gl. Dok 1992- 93.

Exhibitions

Tapetkonk. 1967-68, Kunstindustrimus. 1968; Minitextil, Holstebro Kunstmus. m.fl. 1972; Da. Kunsthåndv. Landssammenslutn., 1976, 1977, 1981, 1982; Kunsthåndv.rådet, Kunstindustrimus. 1978, 1979; KP 1981, 1982; Kunstnernes Sommerudst. 1982, 1984; Rumfang 1988, 1990; Ild og uld, Munkeruphus 1990. Separatudstillinger: Damhuset, Kgs. Lyngby 1980; Farum Kunstforen., 1990; Gal. To'eren, Åkirkeby 1991; endv. i flere firmakunstforen. 1975- 93.

Artworks

Billedvævninger: Livsglæde (1978, Sparekassen SDS, Værløse); Sjælland Rundt, to vævninger af en serie (1979, Lyngby-Taarbæk Rådhus); Blå og grønne toner (1979, Teknologisk Inst., Glostrup); Havnefront (1982, rådhuset i Nuuk); Blågrønne tanker (1982, Nordvang Statshosp., Glostrup); Klippeformationer (1982, plejehj. Runavik, Færøerne); Nykøbing F set fra Lolland, textil rumdeler (1983, Sparekassen SDS, Nyk. F); Billedvævning (1984, Sparekassen SDS, Jyllinge); Livstræ, (1993, Lindevang K., Fr.berg, udf.

K-I3


KUNSTINDEKS DANMARK
& WEILBACHS KUNSTNERLEKSIKON
KUL TUR ARV
FRONT PAGE SEARCH ABOUT HELP LINKS LOGIN DANSK
SEARCH FOR ARTISTS SEARCH FOR ARTWORKS MUSEUMS ARTISTS
AROS - Aarhus Kunstmuseum
Back
Print
Address: Aros Allé 2, 8000 Århus
Phone: 87306600
E-mail: enp@aros.dk
Home page: www.aros.dk
Værker [5765]
Kunstnere [1123]
Contact | Technical and web accessibility | Privacy policy | Rules for use
Kulturarvsstyrelsen: H.C. Andersens Boulevard 2, 1553 København V

K-I4


KUNSTINDEKS DANMARK
& WEILBACHS KUNSTNERLEKSIKON
KUL TUR ARV
FRONT PAGE SEARCH ABOUT HELP LINKS LOGIN DANSK
SEARCH FOR ARTISTS SEARCH FOR ARTWORKS MUSEUMS ARTISTS
KID
uden titel
Back
Print
Literature
Exhibitions
Artist: Abell, Karen
Title: uden titel
Date: 2007
Type of work: Etching
Materials/technique: Etching
Net size: 490 x 392 mm
Signature: Med blyant fntv. E.A. med blyant fnth.. Karen Abell 2007
Museum: Vejle Kunstmuseum, inv. nr. 2010/125
Acquisition: Gave, 28-09-2010
Contact | Technical and web accessibility | Privacy policy | Rules for use
Kulturarvsstyrelsen: H.C. Andersens Boulevard 2, 1553 København V

K-O1

KUNSTINDEKS DANMARK
& WEILBACHS KUNSTNERLEKSIKON

FRONT PAGE | SEARCH | ABOUT | HELP | LINKS | LOGIN | DANSK

SEARCH FOR ARTISTS | SEARCH FOR ARTWORKS | MUSEUMS | ARTISTS

KID

Portræt af Moses og Hanne Ruben

Back
Print

Literature
Exhibitions

Artist: Saloman, Geskel
Title: Portrait of Moses and Hanne Ruben
Date: 1847
Type of work: Painting
Materials/technique: Oil on canvas
Net size: 42 x 51 cm
Museum: Dansk Jødisk Museum, inv. nr. 0213M0027
Acquisition: Købt, 24-04-2012
Person portrayed: Ruben, Moses Magnus (Moshe ben Menachem fra Helsingør) ; Ruben, Hanne (Hendele) f. Eichel ; Ruben, Magnus Ephraim (Menachem Man fra Helsingør) ; Ruben, Hanne (Hanna) f. Salomonsen
Related artworks:

Contact | Technical and web accessibility | Privacy policy | Rules for use

Kulturarvsstyrelsen: H.C. Andersens Boulevard 2, 1553 København V

K-O2

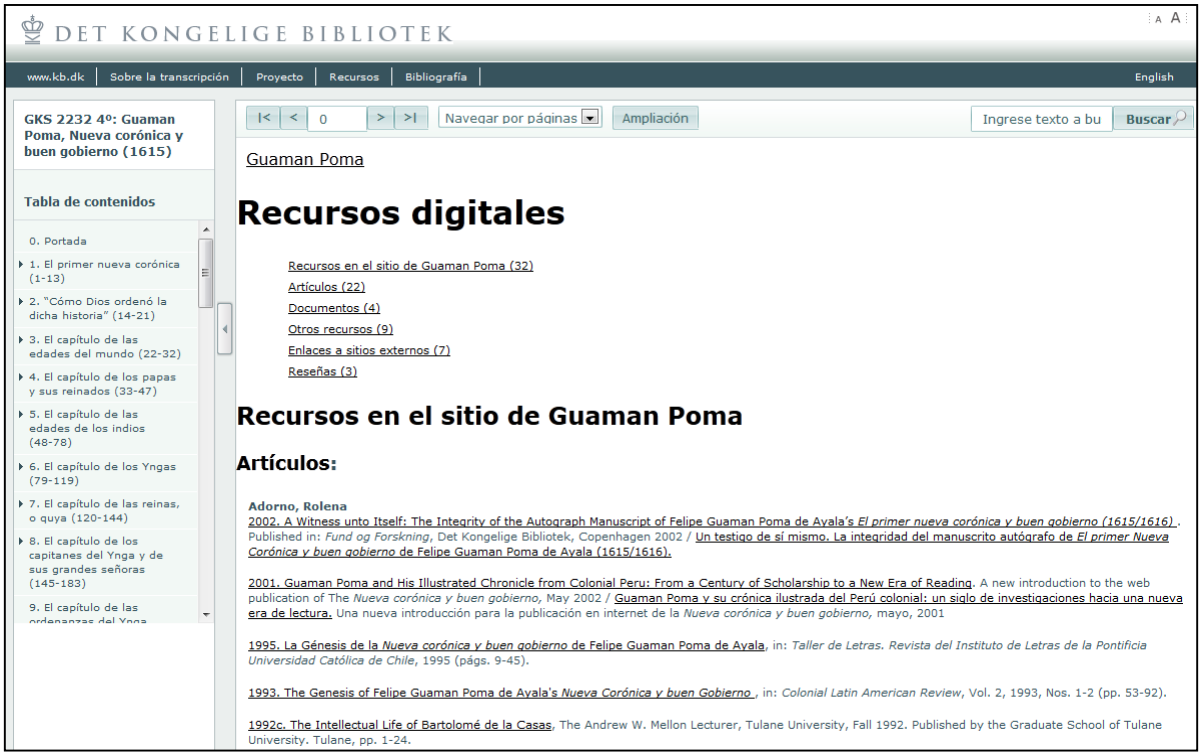

KUNSTINDEKS DANMARK
& WEILBACHS KUNSTNERLEKSIKON
KUL TUR ARV
FRONT PAGE
SEARCH
ABOUT
HELP
LINKS
LOGIN
DANSK
SEARCH FOR ARTISTS
SEARCH FOR ARTWORKS
MUSEUMS
ARTISTS
KID
Vinterlandskab fra Langeland
Back
Print
Literature
Exhibitions
Artist: Fabritius de Tengnagel, F.M.E.
Title: Vinterlandskab fra Langeland
Date: 1828
Type of work: Painting
Materials/technique: Oil on canvas
Net size: 69 x 97,3 cm
Museum: Akademiraadet. Det Kongelige Akademi for de Skønne Kunster, inv. nr. KS 58
Acquisition: 1828
Contact | Technical and web accessibility | Privacy policy | Rules for use
Kulturarvsstyrelsen: H.C. Andersens Boulevard 2, 1553 København V

K-O3

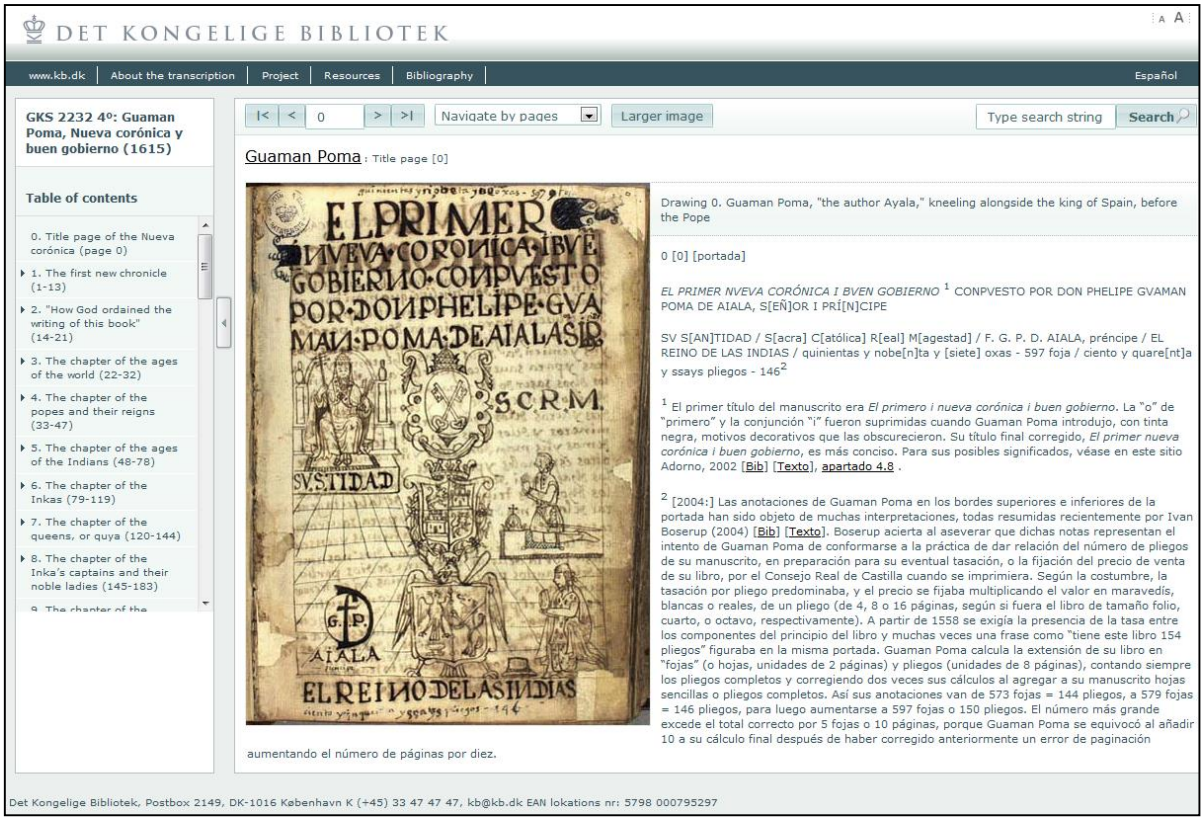
Appendix 3 Screen pictures of studied Poma-objects

Examples of the studied objects, not all objects are included below. Objects with an odd number are English versions of the object and even numbered objects Spanish versions, e.g. P-N1 and P-N2 are different language versions of the same content.

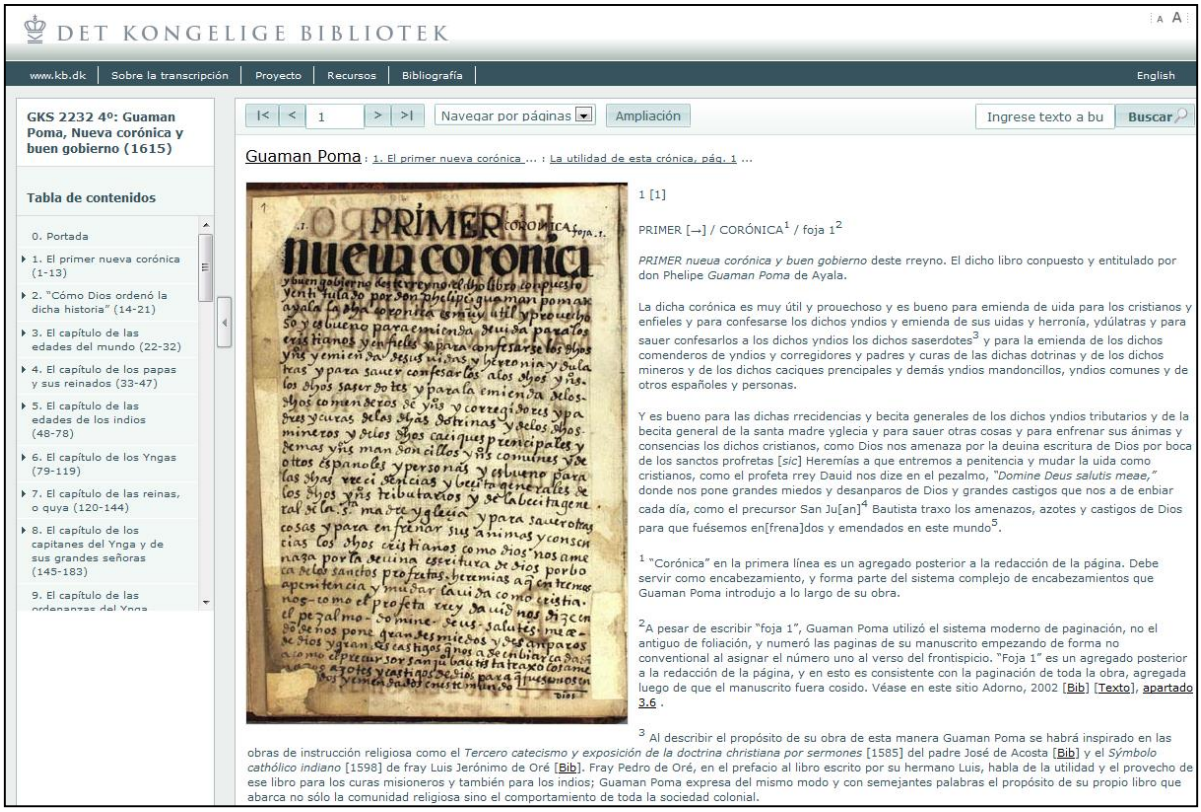
P-N1


DET KONGELIGE BIBLIOTEK
www.kb.dk
Sobre la transcripción
Proyecto
Recursos
Bibliografía
English
GKS 2232 4º: Guaman Poma, Nueva corónica y buen gobierno (1615)
Tabla de contenidos
0. Portada
1. El primer nueva corónica (1-13)
2. "Cómo Dios ordenó la dicha historia" (14-21)
3. El capítulo de las edades del mundo (22-32)
4. El capítulo de los papas y sus reinados (33-47)
5. El capítulo de las edades de los indios (48-78)
6. El capítulo de los Yngas (79-119)
7. El capítulo de las reinas, o quya (120-144)
8. El capítulo de los capitanes del Ynga y de sus grandes señoras (145-183)
9. El capítulo de las ordenanzas del Ynga
Navegar por páginas
Ampliación
Ingrese texto a bu
Buscar
Guaman Poma
Recursos digitales
Recursos en el sitio de Guaman Poma (32)
Artículos (22)
Documentos (4)
Otros recursos (9)
Enlaces a sitios externos (7)
Reseñas (3)
Recursos en el sitio de Guaman Poma
Artículos:
Adorno, Rolena
2002. A Witness unto Itself: The Integrity of the Autograph Manuscript of Felipe Guaman Poma de Ayala's El primer nueva corónica y buen gobierno (1615/1616) . Published in: Fund og Forskning, Det Kongelige Bibliotek, Copenhagen 2002 / Un testigo de sí mismo. La integridad del manuscrito autógrafo de El primer Nueva Corónica y buen gobierno de Felipe Guaman Poma de Ayala (1615/1616).
2001. Guaman Poma and His Illustrated Chronicle from Colonial Peru: From a Century of Scholarship to a New Era of Reading. A new introduction to the web publication of The Nueva corónica y buen gobierno, May 2002 / Guaman Poma y su crónica ilustrada del Perú colonial: un siglo de investigaciones hacia una nueva era de lectura. Una nueva introducción para la publicación en internet de la Nueva corónica y buen gobierno, mayo, 2001
1995. La Génesis de la Nueva corónica y buen gobierno de Felipe Guaman Poma de Ayala, in: Taller de Letras. Revista del Instituto de Letras de la Pontificia Universidad Católica de Chile, 1995 (págs. 9-45).
1993. The Genesis of Felipe Guaman Poma de Ayala's Nueva Corónica y buen Gobierno , in: Colonial Latin American Review, Vol. 2, 1993, Nos. 1-2 (pp. 53-92).
1992c. The Intellectual Life of Bartolomé de la Casas, The Andrew W. Mellon Lecturer, Tulane University, Fall 1992. Published by the Graduate School of Tulane University. Tulane, pp. 1-24.

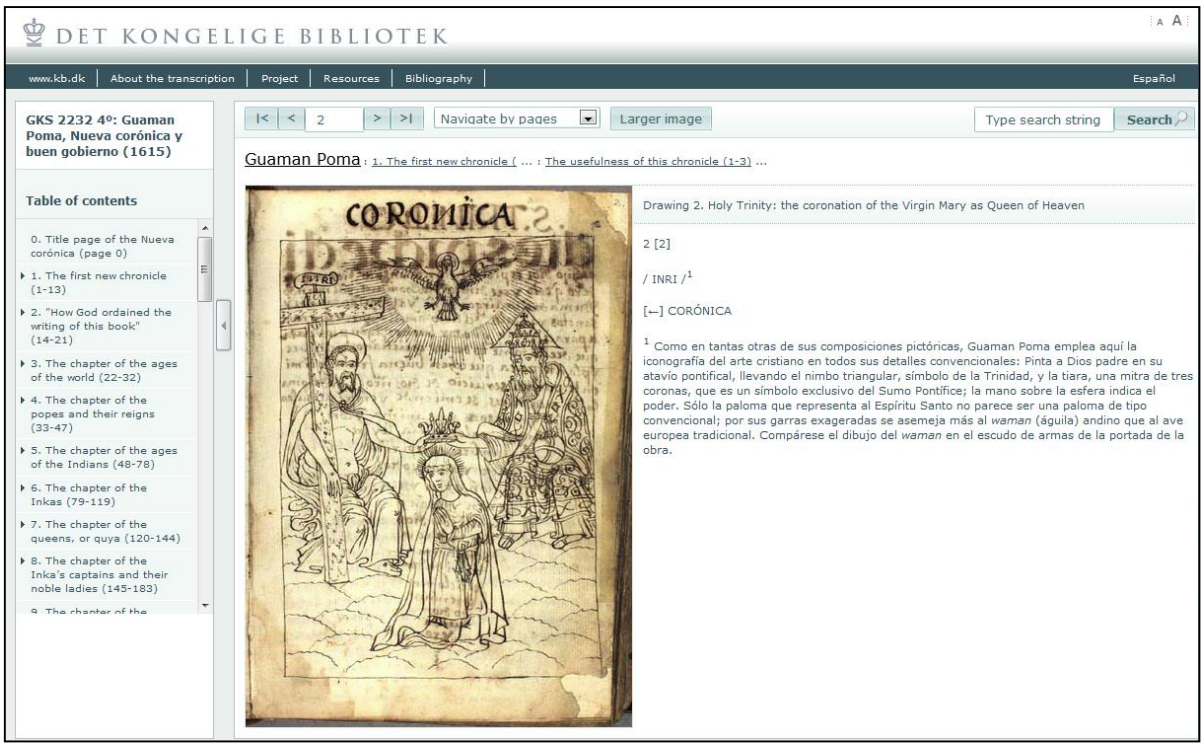
P-N4


DET KONGELIGE BIBLIOTEK
www.kb.dk
About the transcription
Project
Resources
Bibliography
Español
GKS 2232 4º: Guaman Poma, Nueva corónica y buen gobierno (1615)
Table of contents
0. Title page of the Nueva corónica (page 0)
1. The first new chronicle (1-13)
2. "How God ordained the writing of this book" (14-21)
3. The chapter of the ages of the world (22-32)
4. The chapter of the popes and their reigns (33-47)
5. The chapter of the ages of the Indians (48-78)
6. The chapter of the Inkas (79-119)
7. The chapter of the queens, or quya (120-144)
8. The chapter of the Inka's captains and their noble ladies (145-183)
9. The chapter of the
Navigate by pages
Larger image
Type search string
Search
Guaman Poma : Title page [0]
EL PRIMER
NVEVA·CORONICA·I BVE
GOBIERNO·COMPVESTO
POR·DON PHELIPE·GVA
MAN·POMA·DE AIALA S
S.C.R.M.
SVS.TIDAD
G. P.
AIALA
EL REINO DE LAS INDIAS
Drawing 0. Guaman Poma, "the author Ayala," kneeling alongside the king of Spain, before the Pope
0 [0] [portada]
EL PRIMER NVEVA CORÓNICA I BVEN GOBIERNO [1] CONPVESTO POR DON PHELIPE GVAMAN POMA DE AIALA, S[EÑ]OR I PRÍ[N]CIPE
SV S[AN]TIDAD / S[acra] C[atólica] R[eal] M[agestad] / F. G. P. D. AIALA, préncipe / EL REINO DE LAS INDIAS / quinientas y nobe[n]ta y [siete] oxas - 597 foja / ciento y quare[nt]a y ssays pliegos - 146[2]
[1] El primer título del manuscrito era El primero i nueva corónica i buen gobierno. La "o" de "primero" y la conjunción "i" fueron suprimidas cuando Guaman Poma introdujo, con tinta negra, motivos decorativos que las obscurecieron. Su título final corregido, El primer nueva corónica i buen gobierno, es más conciso. Para sus posibles significados, véase en este sitio Adorno, 2002 [Bib] [Texto], apartado 4.8 .
[2] [2004:] Las anotaciones de Guaman Poma en los bordes superiores e inferiores de la portada han sido objeto de muchas interpretaciones, todas resumidas recientemente por Ivan Boserup (2004) [Bib] [Texto]. Boserup acierta al aseverar que dichas notas representan el intento de Guaman Poma de conformarse a la práctica de dar relación del número de pliegos de su manuscrito, en preparación para su eventual tasación, o la fijación del precio de venta de su libro, por el Consejo Real de Castilla cuando se imprimiera. Según la costumbre, la tasación por pliego predominaba, y el precio se fijaba multiplicando el valor en maravedís, blancas o reales, de un pliego (de 4, 8 o 16 páginas, según si fuera el libro de tamaño folio, cuarto, o octavo, respectivamente). A partir de 1558 se exigía la presencia de la tasa entre los componentes del principio del libro y muchas veces una frase como "tiene este libro 154 pliegos" figuraba en la misma portada. Guaman Poma calcula la extensión de su libro en "fojas" (o hojas, unidades de 2 páginas) y pliegos (unidades de 8 páginas), contando siempre los pliegos completos y corregiendo dos veces sus cálculos al agregar a su manuscrito hojas sencillas o pliegos completos. Así sus anotaciones van de 573 fojas = 144 pliegos, a 579 fojas = 146 pliegos, para luego aumentarse a 597 fojas o 150 pliegos. El número más grande excede el total correcto por 5 fojas o 10 páginas, porque Guaman Poma se equivocó al añadir 10 a su cálculo final después de haber corregido anteriormente un error de paginación
aumentando el número de páginas por diez.
Det Kongelige Bibliotek, Postbox 2149, DK-1016 København K (+45) 33 47 47 47, kb@kb.dk EAN lokations nr: 5798 000795297

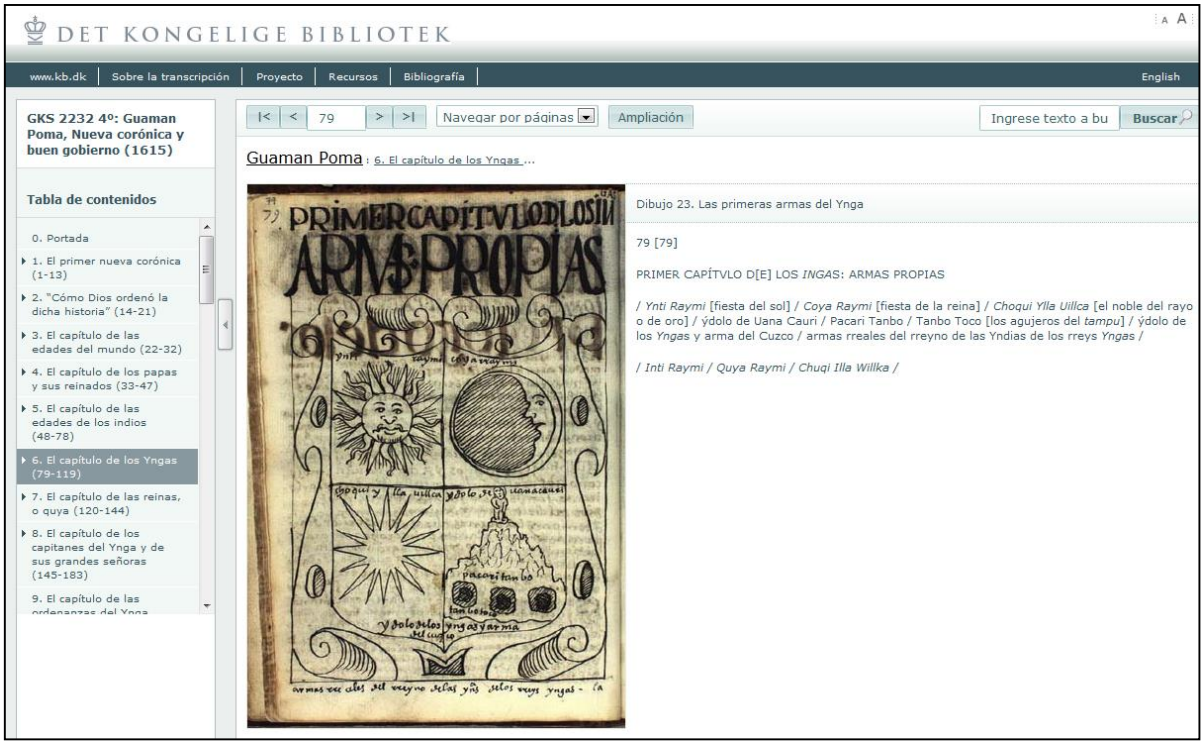
P-O1


DET KONGELIGE BIBLIOTEK
www.kb.dk | Sobre la transcripción | Proyecto | Recursos | Bibliografía | English
GKS 2232 4º: Guaman Poma, Nueva corónica y buen gobierno (1615)
Tabla de contenidos
0. Portada
1. El primer nueva corónica (1-13)
2. "Cómo Dios ordenó la dicha historia" (14-21)
3. El capítulo de las edades del mundo (22-32)
4. El capítulo de los papas y sus reinados (33-47)
5. El capítulo de las edades de los indios (48-78)
6. El capítulo de los Yngas (79-119)
7. El capítulo de las reinas, o quya (120-144)
8. El capítulo de los capitanes del Ynga y de sus grandes señoras (145-183)
9. El capítulo de las ordenanzas del Ynga
Navegar por páginas
Ampliación
Ingrese texto a bu
Buscar
Guaman Poma : 1. El primer nueva corónica ... : La utilidad de esta crónica, pág. 1 ...
PRIMER CORÓNICA foja. 1.
nueua coronica
1 [1]
PRIMER [→] / CORÓNICA[1] / foja 1[2]
PRIMER nueua corónica y buen gobierno deste rreyno. El dicho libro conpuesto y entitulado por don Phelipe Guaman Poma de Ayala.
La dicha corónica es muy útil y prouechoso y es bueno para emienda de uida para los cristianos y enfieles y para confesarse los dichos yndios y emienda de sus uidas y herronía, ydúlatras y para sauer confesarlos a los dichos yndios los dichos saserdotes[3] y para la emienda de los dichos comenderos de yndios y corregidores y padres y curas de las dichas dotrinas y de los dichos mineros y de los dichos caciques prencipales y demás yndios mandoncillos, yndios comunes y de otros españoles y personas.
Y es bueno para las dichas rrecidencias y becita generales de los dichos yndios tributarios y de la becita general de la santa madre yglecia y para sauer otras cosas y para enfrenar sus ánimas y consencias los dichos cristianos, como Dios nos amenaza por la deuina escritura de Dios por boca de los sanctos profretas [sic] Heremías a que entremos a penitencia y mudar la uida como cristianos, como el profeta rrey Dauid nos dize en el pezalmo, "Domine Deus salutis meae," donde nos pone grandes miedos y desanparos de Dios y grandes castigos que nos a de enbiar cada día, como el precursor San Ju[an][4] Bautista traxo los amenazos, azotes y castigos de Dios para que fuésemos en[frena]dos y emendados en este mundo[5].
[1] "Corónica" en la primera línea es un agregado posterior a la redacción de la página. Debe servir como encabezamiento, y forma parte del sistema complejo de encabezamientos que Guaman Poma introdujo a lo largo de su obra.
[2] A pesar de escribir "foja 1", Guaman Poma utilizó el sistema moderno de paginación, no el antiguo de foliación, y numeró las paginas de su manuscrito empezando de forma no convencional al asignar el número uno al verso del frontispicio. "Foja 1" es un agregado posterior a la redacción de la página, y en esto es consistente con la paginación de toda la obra, agregada luego de que el manuscrito fuera cosido. Véase en este sitio Adorno, 2002 [Bib] [Texto], apartado 3.6 .
[3] Al describir el propósito de su obra de esta manera Guaman Poma se habrá inspirado en las obras de instrucción religiosa como el Tercero catecismo y exposición de la doctrina christiana por sermones [1585] del padre José de Acosta [Bib] y el Sýmbolo cathólico indiano [1598] de fray Luis Jerónimo de Oré [Bib]. Fray Pedro de Oré, en el prefacio al libro escrito por su hermano Luis, habla de la utilidad y el provecho de ese libro para los curas misioneros y también para los indios; Guaman Poma expresa del mismo modo y con semejantes palabras el propósito de su propio libro que abarca no sólo la comunidad religiosa sino el comportamiento de toda la sociedad colonial.

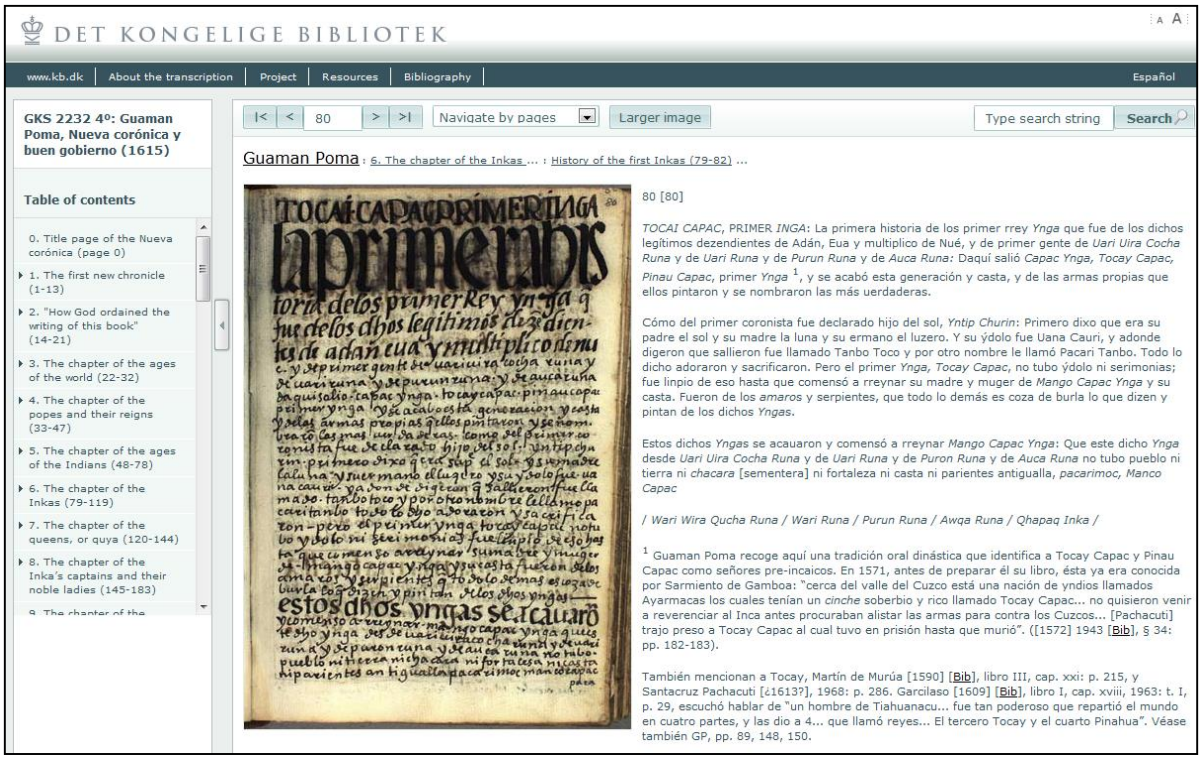
P-O4


DET KONGELIGE BIBLIOTEK
www.kb.dk | About the transcription | Project | Resources | Bibliography | Español
GKS 2232 4º: Guaman Poma, Nueva corónica y buen gobierno (1615)
Table of contents
0. Title page of the Nueva corónica (page 0)
1. The first new chronicle (1-13)
2. "How God ordained the writing of this book" (14-21)
3. The chapter of the ages of the world (22-32)
4. The chapter of the popes and their reigns (33-47)
5. The chapter of the ages of the Indians (48-78)
6. The chapter of the Inkas (79-119)
7. The chapter of the queens, or quya (120-144)
8. The chapter of the Inka's captains and their noble ladies (145-183)
9. The chapter of the
Navigate by pages
Larger image
Type search string
Search
Guaman Poma : 1. The first new chronicle (... : The usefulness of this chronicle (1-3) ...
CORONICA
Drawing 2. Holy Trinity: the coronation of the Virgin Mary as Queen of Heaven
2 [2]
/ INRI /[1]
[←] CORÓNICA
[1] Como en tantas otras de sus composiciones pictóricas, Guaman Poma emplea aquí la iconografía del arte cristiano en todos sus detalles convencionales: Pinta a Dios padre en su atavío pontifical, llevando el nimbo triangular, símbolo de la Trinidad, y la tiara, una mitra de tres coronas, que es un símbolo exclusivo del Sumo Pontífice; la mano sobre la esfera indica el poder. Sólo la paloma que representa al Espíritu Santo no parece ser una paloma de tipo convencional; por sus garras exageradas se asemeja más al waman (águila) andino que al ave europea tradicional. Compárese el dibujo del waman en el escudo de armas de la portada de la obra.

P-O5


DET KONGELIGE BIBLIOTEK
www.kb.dk | Sobre la transcripción | Proyecto | Recursos | Bibliografía | English
GKS 2232 4º: Guaman Poma, Nueva corónica y buen gobierno (1615)
Tabla de contenidos
0. Portada
1. El primer nueva corónica (1-13)
2. "Cómo Dios ordenó la dicha historia" (14-21)
3. El capítulo de las edades del mundo (22-32)
4. El capítulo de los papas y sus reinados (33-47)
5. El capítulo de las edades de los indios (48-78)
6. El capítulo de los Yngas (79-119)
7. El capítulo de las reinas, o quya (120-144)
8. El capítulo de los capitanes del Ynga y de sus grandes señoras (145-183)
9. El capítulo de las ordenanzas del Ynga
Navegar por páginas
Ampliación
Ingrese texto a bu
Buscar
Guaman Poma : 6. El capítulo de los Yngas ...
PRIMER CAPÍTVLO D[E] LOS INGAS: ARMAS PROPIAS
Dibujo 23. Las primeras armas del Ynga
79 [79]
PRIMER CAPÍTVLO D[E] LOS INGAS: ARMAS PROPIAS
/ Ynti Raymi [fiesta del sol] / Coya Raymi [fiesta de la reina] / Choqui Ylla Uillca [el noble del rayo o de oro] / ýdolo de Uana Cauri / Pacari Tanbo / Tanbo Toco [los agujeros del tampu] / ýdolo de los Yngas y arma del Cuzco / armas rreales del rreyno de las Yndias de los rreys Yngas /
/ Inti Raymi / Quya Raymi / Chuqi Illa Willka /

P-O8

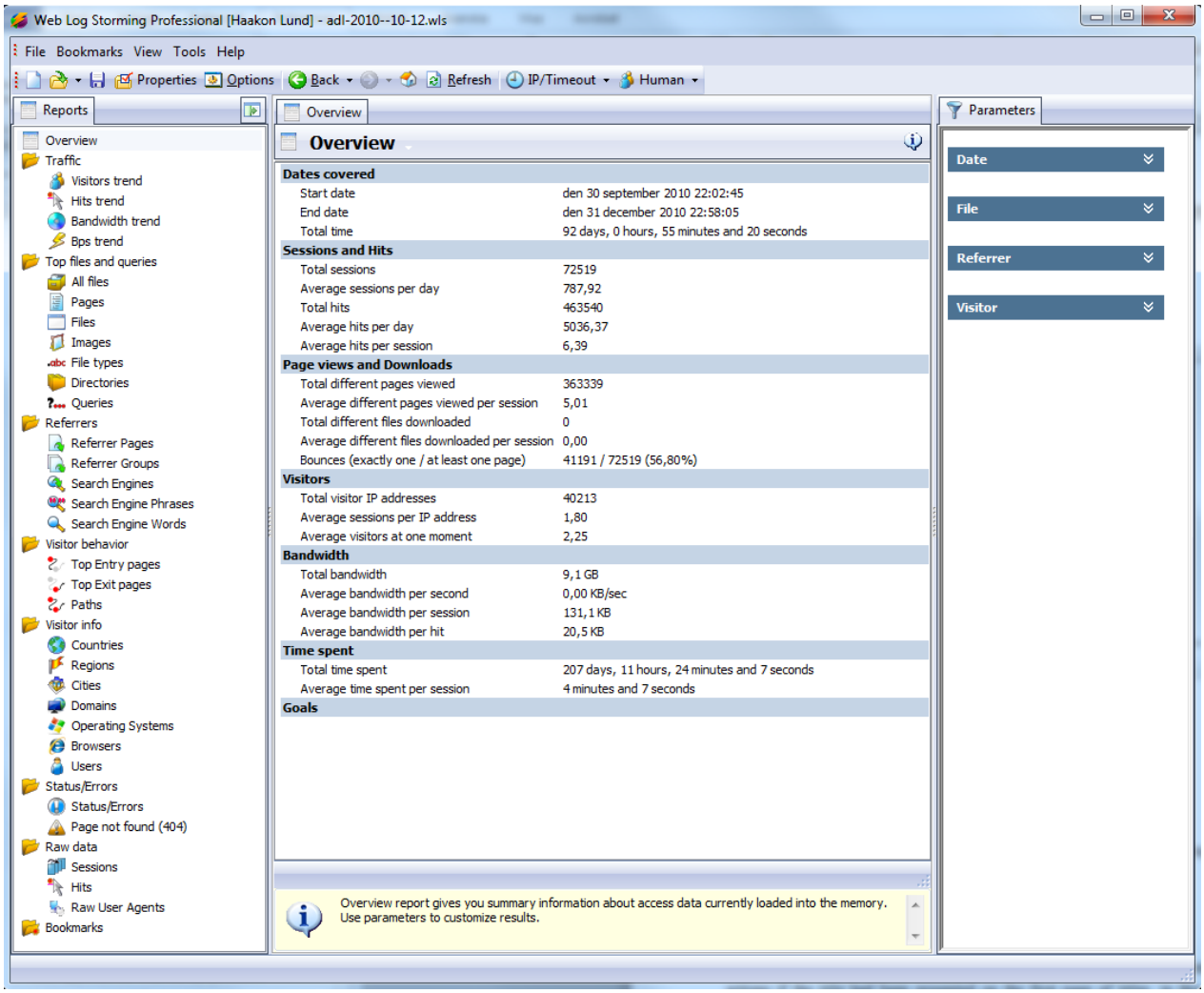

DET KONGELIGE BIBLIOTEK
www.kb.dk | About the transcription | Project | Resources | Bibliography | Español
GKS 2232 4º: Guaman Poma, Nueva corónica y buen gobierno (1615)
Table of contents
0. Title page of the Nueva corónica (page 0)
1. The first new chronicle (1-13)
2. "How God ordained the writing of this book" (14-21)
3. The chapter of the ages of the world (22-32)
4. The chapter of the popes and their reigns (33-47)
5. The chapter of the ages of the Indians (48-78)
6. The chapter of the Inkas (79-119)
7. The chapter of the queens, or quya (120-144)
8. The chapter of the Inka's captains and their noble ladies (145-183)
9. The chapter of the
Navigate by pages
Larger image
Type search string
Search
Guaman Poma : 6. The chapter of the Inkas ... : History of the first Inkas (79-82) ...
80 [80]
TOCAI CAPAC, PRIMER INGA: La primera historia de los primer rrey Ynga que fue de los dichos legítimos dezendientes de Adán, Eua y multiplico de Nué, y de primer gente de Uari Uira Cocha Runa y de Uari Runa y de Purun Runa y de Auca Runa: Daquí salió Capac Ynga, Tocay Capac, Pinau Capac, primer Ynga [1], y se acabó esta generación y casta, y de las armas propias que ellos pintaron y se nombraron las más uerdaderas.
Cómo del primer coronista fue declarado hijo del sol, Yntip Churin: Primero dixo que era su padre el sol y su madre la luna y su ermano el luzero. Y su ýdolo fue Uana Cauri, y adonde digeron que sallieron fue llamado Tanbo Toco y por otro nombre le llamó Pacari Tanbo. Todo lo dicho adoraron y sacrificaron. Pero el primer Ynga, Tocay Capac, no tubo ýdolo ni serimonias; fue linpio de eso hasta que comensó a rreynar su madre y muger de Mango Capac Ynga y su casta. Fueron de los amaros y serpientes, que todo lo demás es coza de burla lo que dizen y pintan de los dichos Yngas.
Estos dichos Yngas se acauaron y comensó a rreynar Mango Capac Ynga: Que este dicho Ynga desde Uari Uira Cocha Runa y de Uari Runa y de Puron Runa y de Auca Runa no tubo pueblo ni tierra ni chacara [sementera] ni fortaleza ni casta ni parientes antigualla, pacarimoc, Manco Capac
/ Wari Wira Qucha Runa / Wari Runa / Purun Runa / Awqa Runa / Qhapaq Inka /
[1] Guaman Poma recoge aquí una tradición oral dinástica que identifica a Tocay Capac y Pinau Capac como señores pre-incaicos. En 1571, antes de preparar él su libro, ésta ya era conocida por Sarmiento de Gamboa: "cerca del valle del Cuzco está una nación de yndios llamados Ayarmacas los cuales tenían un cinche soberbio y rico llamado Tocay Capac... no quisieron venir a reverenciar al Inca antes procuraban alistar las armas para contra los Cuzcos... [Pachacuti] trajo preso a Tocay Capac al cual tuvo en prisión hasta que murió". ([1572] 1943 [Bib], § 34: pp. 182-183).
También mencionan a Tocay, Martín de Murúa [1590] [Bib], libro III, cap. xxi: p. 215, y Santacruz Pachacuti [¿1613?], 1968: p. 286. Garcilaso [1609] [Bib], libro I, cap. xviii, 1963: t. I, p. 29, escuchó hablar de "un hombre de Tiahuanacu... fue tan poderoso que repartió el mundo en cuatro partes, y las dio a 4... que llamó reyes... El tercero Tocay y el cuarto Pinahua". Véase también GP, pp. 89, 148, 150.

P-O9

Appendix 4 Example of a session in ADL

Time	Requested URL	Referrer
2010-09-30 23:01	www.adl.dk/adl_pub/forfatter/e_forfatter/e_forfatter.xsql? ff_id=75	www.google.dk/url?sa=t&source=web&cd=5&ved=0CDAQFjAE&url=http%3A%2F%2Fwww.adl.dk%2Fadl_pub%2Fforfatter%2Fe_forfatter%2Fe_forfatter.xsql%3Fff_id%3D75&rct=j&q=kierkegaard&ei=eS6lTP6xB8TMswa57Z2lCA&usg=[+session id]
+00:00:06	www.adl.dk/adl_pub/udgave/udgave_menu.xsql?ff_id=75&nnoc=adl_pub	www.adl.dk/adl_pub/forfatter/e_forfatter/e_forfatter.xsql?ff_id=75
+00:00:18	www.adl.dk/adl_pub/vaerker/cv/ff_vaerker_menu.xsql?ff_id=75&bogstav=&nnoc=adl_pub	www.adl.dk/adl_pub/udgave/udgave_menu.xsql?ff_id=75&nnoc=adl_pub
+00:00:25	www.adl.dk/adl_pub/periode/cv/periode_menu.xsql? nnoc=adl_pub	www.adl.dk/adl_pub/vaerker/cv/ff_vaerker_menu.xsql?ff_id=75&bogstav=&nnoc=adl_pub
+00:00:38	www.adl.dk/adl_pub/forfatter/e_forfatter/e_forfatter.xsql?nnoc=adl_pub&ff_id=75	www.adl.dk/adl_pub/periode/cv/periode_menu.xsql?nnoc=adl_pub
+00:00:50	www.adl.dk/adl_pub/fportraet/cv/ShowFpItem.xsql? ff_id=75&p_fpkat_id=indl&nnoc=adl_pub	www.adl.dk/adl_pub/forfatter/e_forfatter/e_forfatter.xsql?nnoc=adl_pub&ff_id=75
+00:00:56	www.adl.dk/adl_pub/fportraet/cv/ShowFpItem.xsql?nnoc=adl_pub&ff_id=75&p_fpkat_id=bibliog	www.adl.dk/adl_pub/fportraet/cv/ShowFpItem.xsql?ff_id=75&p_fpkat_id=indl&nnoc=adl_pub
+00:01:26	www.adl.dk/adl_pub/fportraet/cv/ShowFpItem.xsql?ff_id=75&p_fpkat_id=indl&nnoc=adl_pub	www.adl.dk/adl_pub/forfatter/e_forfatter/e_forfatter.xsql?nnoc=adl_pub&ff_id=75
+00:01:27	www.adl.dk/adl_pub/forfatter/e_forfatter/e_forfatter.xsql?nnoc=adl_pub&ff_id=75	www.adl.dk/adl_pub/periode/cv/periode_menu.xsql?nnoc=adl_pub
+00:01:28	www.adl.dk/adl_pub/vaerker/cv/ff_vaerker_menu.xsql?ff_id=75&bogstav=&nnoc=adl_pub	www.adl.dk/adl_pub/udgave/udgave_menu.xsql?ff_id=75&nnoc=adl_pub
	www.adl.dk/adl_pub/periode/cv/periode_menu.xsql?nnoc=adl_pub	www.adl.dk/adl_pub/vaerker/cv/ff_vaerker_menu.xsql?ff_id=75&bogstav=&nnoc=adl_pub
	www.adl.dk/adl_pub/udgave/udgave_menu.xsql?ff_id=75&nnoc=adl_pub	www.adl.dk/adl_pub/forfatter/e_forfatter/e_forfatter.xsql?ff_id=75
+00:01:30	www.adl.dk/adl_pub/forfatter/e_forfatter/e_forfatter.xsql?ff_id=75	www.google.dk/url?sa=t&source=web&cd=5&ved=0CDAQFjAE&url=http%3A%2F%2Fwww.adl.dk%2Fadl_pub%2Fforfatter%2Fe_forfatter%2Fe_forfatter.xsql%3Fff_id%3D75&rct=j&q=kierkegaard&ei=eS6lTP6xB8TMswa57Z2lCA&usg=[+session id]

Appendix 5 Screen shot of Web Log Storming



Overview screen in Web Log Storming (the ADL logs are loaded).

Appendix 6 Data about the sorting based on referrer in WLS

Referring search engines with a share >0.05% in WLS.

	ADL	KID	Poma
1	Google	google	google
2	Bing	bing	bing
3	find.tdc	webcache.googleusercontent	images.google
4	webcache.googleusercontent	translate.googleusercontent	webcache.googleusercontent
5	dk.search.yahoo	search.yahoo	search.yahoo
6	translate.googleusercontent	find.tdc	translate.googleusercontent
7		dk.search.yahoo	translate.google
8			images.search. yahoo
9			mx.search.yahoo
			ar.search.yahoo

The strings used in WLS when sorting the search engine sessions.

	Search string
ADL	google,bing,find.tdc,webcache.googleusercontent,dk.search.yahoo,translate.googleusercontent
KID	google,bing,webcache.googleusercontent,translate.googleusercontent,search.yahoo,find.tdc,dk.search.yahoo
Poma	google,bing,images.google,webcache.googleusercontent,search.yahoo,translate.googleusercontent,translate.google,images.search.yahoo,mx.search.yahoo,ar.search.yahoo

The strings used in WLS when sorting the link sessions.

	Search string
ADL	-*direct*,-google,-bing,-find.tdc,-webcache.googleusercontent,-dk.search.yahoo,-translate.googleusercontent
KID	-*direct*,-google,-bing,-webcache.googleusercontent,-translate.googleusercontent,-search.yahoo,-find.tdc,-dk.search.yahoo
Poma	-*direct*,-google,-bing,-images.google,-webcache.googleusercontent,-search.yahoo,-translate.googleusercontent,-translate.google,-images.search.yahoo,-mx.search.yahoo,-ar.search.yahoo

Appendix 7 Web survey questions

Survey questions English

1. How did you reach this site?
    a. Direct via a bookmark.
    b. Direct by typing in the url.
    c. Through links on the web (web pages, blogs, etc.)
    d. By using a search engine for a topical search (e.g. author or artists name as search terms).
    e. By using a search engine to find known site (e.g. parts of title or url as search terms).

2. How often have you visited this resource?
    a. This is the first time.
    b. I have visited the web site a couple of times (1-5 times).
    c. I have often visited the web site (more than 5 times).
    d. I have often visited the web site (more than 5 times).

3. Why are you visiting this resource today?
    a. Exploring the site and its content.
    b. Learning about a topic.
    c. Look up fact.
    d. Just curious.
    e. Other -- please specify:

4. In what context do you visit the web site?
    a. School or study visit.
    b. Hobby or leisure.
    c. Work.
    d. By coincidence.
    e. Other -- please specify:

5. How do you rate your knowledge about the Internet?
    a. Excellent.
    b. Good.
    c. Could be better.
    d. Bad.

6. How do you rate your skills in using Internet and Web technologies, e.g. using web browsers, web search engines and other web tools?
    a. Excellent.
    b. Good.
    c. Could be better.
    d. Bad.

7. How do you rate your ability to evaluate information on the Web with regard to its relevance, quality and credibility?
    a. Excellent.
    b. Good.
    c. Could be better.
    d. Bad.

8. Your age?
    a. -18
    b. 19—30
    c. 31—45
    d. 46—65
    e. 66-

9. Your gender?
    a. Female
    b. Male

10. Where do you live?
    a. Denmark.
    b. Other Scandinavian country.
    c. Other European country.
    d. North America.
    e. South America.
    f. Austraila or New Zeeland.
    g. Asia.
    h. Africa.
11. How many years have you gone to school (in the educational system)? (e.g. Elementry school 8-9 years, Profession education (e.g. Carpenter, clerk) 12-13 years, University 15-20 years)
    a. --6 years
    b. 7—10
    c. 11—13
    d. 14—16
    e. 17—20
    f. 21—

12. What is your present position? (e.g. Pensioned, pupil, student, carpenter, teacher)
    a. Open:

Survey questions Danish

1. Hvordan fandt du frem til dette websted?
    a. Direkte via et bogmærke ("favoritter")
    b. Direkte ved at indtaste web-adressen
    c. Gennem links på nettet (hjemmesider, blogs osv.)
    d. Ved at søge på et emne i en søgemaskine (f.eks. forfatter- eller kunstnernavn som søgeord)
    e. Ved at bruge en søgemaskine til at finde et kendt websted (f.eks. dele af webstedets navn eller adresse som søgeord)

2. Hvor ofte har du besøgt dette websted? (Alt: Hvor ofte har du været inde på dette websted?)
    a. Dette er første gang.
    b. Jeg har besøgt/været inde på webstedet nogle få gange (1-5 gange)
    c. Jeg har ofte besøgt/været inde på webstedet (mere end 5 gange)
    d. Ved ikke

3. Hvorfor besøger du/ er du inde på webstedet i dag? (afkryds gerne flere valg)
    a. For at udforske stedet og dets indhold
    b. For at få viden om et bestemt emne
    c. For at slå oplysninger/fakta op
    d. Af nysgerrighed
    e. Andet - uddyb venligst

4. I hvilken forbindelse/sammenhæng er du inde på webstedet?
    a. Skole eller studier
    b. Hobby eller fritid
    c. Arbejde
    d. Ved et tilfælde
    e. Andet - uddyb venligst

5. Hvordan vil du vurdere din viden om internettet?
    a. Fremragende
    b. God
    c. Kunne være bedre
    d. Dårlig

6. Hvordan vil du vurdere dine færdigheder i brug af internettet og web teknologier, f.eks. brug af browsere, søgemaskiner og andre web-redskaber?
    a. Fremragende
    b. God
    c. Kunne være bedre
    d. Dårlig

7. Hvordan vil du vurdere din evne til at bedømme information på nettet med hensyn til relevans, kvalitet og troværdighed?
    a. Fremragende
    b. God
    c. Kunne være bedre
    d. Dårlig

8. Hvad er din alder?
    a. -18
    b. 19—30
    c. 31—45
    d. 46—65
    e. 66-

9. Hvad er dit køn?
    a. Kvinde
    b. Mand

10. Hvor bor du?
    a. Danmark
    b. Et andet skandinavisk land
    c. Et andet europæisk land
    d. Nordamerika
    e. Sydamerika
    f. Australien eller New Zealand
    g. Asien
    h. Afrika

11. Hvor mange års skolegang har du? (f.eks. folkeskole 8-9 år, erhvervsuddannelse (f.eks. tømrer, kontorassistent) 12-13 år, universitetsuddannelse 15-20 år)
    a. --6 years
    b. 7—10
    c. 11—13
    d. 14—16
    e. 17—20
    f. 21—

12. Hvad er din nuværende beskæftigelse? (f.eks. pensionist, skoleelev, studerende, tømrer, lærer)
    a. Skriv her (ingen afkrydsning):

Survey questions Spanish

1. ¿Cómo llegaste a esta página?
    a. Directamente a través de un marcador/favorites
    b. Tecleando directamente la url.
    c. A través de enlaces en la web (páginas web, blogs, etc.)
    d. Usando un motor de búsqueda para una búsqueda por tema (por ejemplo, autor o nombres de artistas como términos de búsqueda)
    e. Usando un motor de búsqueda para encontrar una página conocida (por ejemplo, parte de un título o una url como términos de búsqueda)

2. ¿Cuántas veces has visitado esta fuente/recurso?
    a. Esta es la primera vez
    b. He visitado la página web unas cuantas veces (1-5 veces)
    c. He visitado a menudo la página (más de 5 veces)
    d. No lo sé

3. ¿Por qué estas visitando esta fuente/recurso hoy? (múltiples respuestas)
    a. Explorando la página y su contenido
    b. Aprendiendo sobre un tema
    c. Buscando algo en concreto
    d. Sólo por curiosidad

4. ¿En qué contexto visitas la página web?
    a. Por motivos académicos o de studios
    b. Hobby u ocio
    c. Trabajo
    d. Por casualidad
    e. Otro - por favor, especifica:

5. ¿Cómo valoras tu conocimiento sobre Internet?
    a. Excelente
    b. Bueno
    c. Podría ser major
    d. Malo

6. ¿Cómo valoras tus habilidades usando Internet y las tecnologías Web? Por ejemplo, al usar navegadores web, motores de búsqueda y otras herramientas web
    a. Excelente
    b. Bueno
    c. Podría ser major
    d. Malo

7. ¿Cómo valoras tu capacidad para evaluar información en la Web respecto a su relevancia, calidad y credibilidad?
    a. Excelente
    b. Bueno
    c. Podría ser major
    d. Malo

8. ¿Edad?
    a. -18
    b. 19—30
    c. 31—45
    d. 46—65
    e. 66-

9. ¿Sexo?
    a. Mujer
    b. Hombre

10. ¿Dónde vives?
    a. Dinamarca
    b. Otro país escandinavo
    c. Otro país europeo
    d. Norteamérica
    e. Sudamérica
    f. Australia o Nueva Zelanda
    g. Asia
    h. África

11. ¿Cuántos años has estudiado (en el sistema educativo)? (p.e. Escuela elemental 8-9 años, Educación profesional (p.e. Carpintero, dependiente) 12-13 años, Universidad 15-20 años)
    a. --6 years
    b. 7—10
    c. 11—13
    d. 14—16
    e. 17—20
    f. 21—

12. ¿Cuál es tu situación actual?
    a. Respuesta abierta:

Appendix 8 Site structure analysis (including URL analysis)

URL analysis: In the case of ADL and KID the URLs has to be modified so the log files can be analysed regarding sessions. The question mark has been replaced with another sign, a minus (-) so the software won't treat the right part of the URLs as questions.

Site structure analysis of the structure and content, including URL analysis, in ADL.

Type	URL	Level
Top	http://adl.dk/adl_pub/forside/cv/forside.xsql ?nnoc=adl_pub	N
Authors	http://adl.dk/adl_pub/forfatter/forfatter_menu.xsql ?nnoc=adl_pub	N
Authors A	http://adl.dk/adl_pub/forfatter/forfatter_menu.xsql?nnoc=adl_pub#A	N
H.C. Andersen top	http://adl.dk/adl_pub/forfatter/e_forfatter/e_forfatter.xsql?ff_id=22 &nnoc=adl_pub	I
H.C. Andersen title list - start	http://adl.dk/adl_pub/vaerker/cv/ff_vaerker_menu.xsql?ff_id=22 &bogstav=&nnoc=adl_pub	I
H.C. Andersen title list - D	http://adl.dk/adl_pub/vaerker/cv/ff_vaerker_menu.xsql?ff_id=22&bogstav=D &nnoc=adl_pub	I
H.C. Andersen: Den lille Havfrue, Se værket i flg. udgivelser:	http://adl.dk/adl_pub/vaerker/cv/e_vaerk/e_vaerk.xsql?ff_id=22&id=2247 &hist=fmD&nnoc=adl_pub	I
H.C. Andersen: Den lille Havfrue, facimile	http://adl.dk/adl_pub/pg/cv/ShowPgImg.xsql?p_udg_id=93&p_sidenr=87 &hist=&nnoc=adl_pub	O
H.C. Andersen: Den lille Havfrue, text	http://adl.dk/adl_pub/pg/cv/ShowPgText.xsql?p_udg_id=93&p_sidenr=87 &hist=&nnoc=adl_pub	O
H.C. Andersen: Den lille Havfrue, downloadable text	http://adl.dk/adl_pub/pg/cv/AsciiPgVaerk2.xsql?nnoc=adl_pub&p_udg_id=93&p_vaerk_id=2247	O
H.C. Andersen: used editions	http://adl.dk/adl_pub/udgave/udgave_menu.xsql?ff_id=22 &nnoc=adl_pub	I
H.C. Andersen manus menu	http://adl.dk/adl_pub/manus/cv/ff_manus_menu.xsql?ff_id=22 &nnoc=adl_pub	I
H.C. Andersen manus: Manuskript til 'H.C. Andersens eventyr'	http://adl.dk/adl_pub/manus/cv/e_manus/e_manus.xsql?ff_id=22&id=47 &vaerk_id=&hist=manus&nnoc=adl_pub	I
H.C. Andersen: notes	http://adl.dk/adl_pub/node/cv/NodeListe.xsql?ff_id=22 &nnoc=adl_pub	I
H.C. Andersen: author portrait	http://adl.dk/adl_pub/fportraet/cv/ShowFpItem.xsql?nnoc=adl_pub&ff_id=22&p_fpkat_id=indl	I
H.C. Andersen: biography (+internal links on page)	http://adl.dk/adl_pub/fportraet/cv/ShowFpItem.xsql?nnoc=adl_pub&ff_id=22&p_fpkat_id=biog	I
H.C. Andersen: about biography author	http://adl.dk/adl_pub/fportraet/cv/ShowFpItem.xsql?nnoc=adl_pub&ff_id=22&p_fpkat_id=port	I

Type	URL	Level
H.C. Andersen: author portrait as pdf	http://adl.dk/adl_pub/fportraet/cv/FpPdf.xsql?nnoc=adl_pub&ff_id=22	I
Period	http://adl.dk/adl_pub/periode/cv/periode_menu.xsql ?nnoc=adl_pub	I
Period: about Romanticism, introduction	http://adl.dk/adl_pub/periode/cv/ShowPerbesItem.xsql?nnoc=adl_pub&p_perbeskat_id=indled&p_periode_id=20	I
Period: about Romanticism, the period	http://adl.dk/adl_pub/periode/cv/ShowPerbesItem.xsql?nnoc=adl_pub&p_periode_id=20&p_perbeskat_id=period	I
Period: about Romanticism, bibliography	http://adl.dk/adl_pub/periode/cv/ShowPerbesItem.xsql?nnoc=adl_pub&p_periode_id=20&p_perbeskat_id=biblio	I
Period: about Romanticism, description author	http://adl.dk/adl_pub/periode/cv/ShowPerbesItem.xsql?nnoc=adl_pub&p_periode_id=20&p_perbeskat_id=forf	I
Title A (A-Ö)	http://adl.dk/adl_pub/vaerker/cv/vaerker_menu.xsql?bogstav=A &nnoc=adl_pub	I
Manus (sorted on title)	http://adl.dk/adl_pub/manus/cv/manus_menu_sm.xsql ?nnoc=adl_pub	I
Manus (sorted on author)	http://adl.dk/adl_pub/manus/cv/manus_menu_sf.xsql ?nnoc=adl_pub	I
List of notes (sorted on title)	http://adl.dk/adl_pub/node/cv/node_menu.xsql ?nnoc=adl_pub	I
Search 1	http://adl.dk/adl_pub/soeg/cv/search_menu.xsql ?nnoc=adl_pub	I
Search 2a title	http://adl.dk/adl_pub/soeg/cv/thematic/ThematicSearch.xsql ?nnoc=adl_pub	I
Search 2b fulltext	http://adl.dk/adl_pub/soeg/cv/fritekst/fritekst_soegning.xsql ?nnoc=adl_pub	I
SERP	http://adl.dk/adl_pub/pg/cv/AdvTextSearchResults.xsql	I

Site structure analysis of the structure and content, including URL analysis, in KID.

Type	URL	Level
Top (incl. Search for artists & Search for artwork)	https://www.kulturarv.dk/kid/Forside.do;jsessionid=7D6ACFA315D9AC4AD0A3F16862D0E94C	N
Advanced search for artists	https://www.kulturarv.dk/kid/SoegKunstner.do	N
Artist search: SERP	https://www.kulturarv.dk/kid/SoegResultatKunstnerRefresh.do?page=1&orderBy=asc:sort_navn&action=SoegResultatKunstnerRefresh&listeviewtype=soegkunstnerliste	N
Artist: Karen Abel	https://www.kulturarv.dk/kid/VisKunstner.do?kunstnerId=8135	I
Artworks by Karen Abel	https://www.kulturarv.dk/kid/SoegKunstnerVaerker.do?kunstnerId=8135	I
Artwork 1 by Karen Abel (uden titel)	https://www.kulturarv.dk/kid/VisVaerk.do?vaerkId=450177	O
Information from Weilbachs Kunstnerleksikon: Karen Abel (all) (categories: genealogy, exhibitions, travels, education, occupations, biography, artworks, scholarships, literature, all)	https://www.kulturarv.dk/kid/VisWeilbach.do?kunstnerId=8135&wsektion=alle	I
Advanced search for artwork	https://www.kulturarv.dk/kid/SoegVaerk.do	N
Artwork search: SERP	https://www.kulturarv.dk/kid/SoegResultatVaerk.do?orderBy=asc:sort_titel&page=0&action=SoegResultatVaerkRefresh&listeviewtype=soegvaerkliste	N
Artwork 1 in SERP	https://www.kulturarv.dk/kid/VisVaerk.do?vaerkId=20523	O
Artists	https://www.kulturarv.dk/kid/SoegKunstneroversigt.do	N
Artists: A	https://www.kulturarv.dk/kid/SoegKunstneroversigtRefresh.do?prefix=A	N
Museums in Kunstindeks Danmark	https://www.kulturarv.dk/kid/SoegMuseumsoversigt.do	N
Museums in Kunstindeks Danmark: A	https://www.kulturarv.dk/kid/SoegMuseumsoversigt.do?prefix=A	N
Museum: Vejle Kunstmuseum	https://www.kulturarv.dk/kid/VisMuseum.do?museumId=233	I
About website	https://www.kulturarv.dk/kid/Websted.do	N
About Kunstindeks Danmark	https://www.kulturarv.dk/kid/OmKID.do	N
About Weilbachs Kunstnerleksikon	https://www.kulturarv.dk/kid/OmWeilbach.do	N
Help: FAQ	https://www.kulturarv.dk/kid/Hjaelp.do	N
Help: Search tips	https://www.kulturarv.dk/kid/Soegetips.do	N
Help: Contact	https://www.kulturarv.dk/kid/Kontakt.do	N
Links	https://www.kulturarv.dk/kid/Links.do	N

Site structure analysis of the structure and content, including URL analysis, in Poma.

Type	URL	Level
Top EN	http://www.kb.dk/permalink/2006/poma/info/en/frontpage.htm	N
About transcription	http://www.kb.dk/permalink/2006/poma/info/en/foreword.htm	N
About the project	http://www.kb.dk/permalink/2006/poma/info/en/project/project.htm	N
Digital resources	http://www.kb.dk/permalink/2006/poma/info/en/docs/index.htm	N
Digital resource – Article – Ossio 1998 (example)	http://www.kb.dk/permalink/2006/poma/info/en/docs/ossio/1998/index.htm	N
Bibliography	http://www.kb.dk/permalink/2006/poma/info/en/biblio/index.htm	N
Title page (Drawing 0 [1])	http://www.kb.dk/permalink/2006/poma/titlepage/en/text/?open=id2971047	O
Page 1	http://www.kb.dk/permalink/2006/poma/1/en/text/?open=id2971082	O
Page 2 (Drawing 2)	http://www.kb.dk/permalink/2006/poma/2/en/text/?open=id2971082	O
Page 3	http://www.kb.dk/permalink/2006/poma/3/en/text/?open=id2971082	O
Page 4	http://www.kb.dk/permalink/2006/poma/4/en/text/?open=id2971082	O
Page 1188	http://www.kb.dk/permalink/2006/poma/1188/en/text/?open=id2979068	O
Page 1189 (Drawing 398)	http://www.kb.dk/permalink/2006/poma/1189/en/image/?open=id2649679	O

Appendix 9 Referring search engines

Referring search engines in ADL (top 10).

Search Engine	#	%
google	41748	98,37%
bing	295	0,70%
webcache.googleusercontent	87	0,21%
dk.search.yahoo	80	0,19%
find.tdc	65	0,15%
translate.googleusercontent	49	0,12%
theeuropeanlibrary	22	0,05%
search.yahoo	17	0,04%
dk.yhs.search.yahoo	17	0,04%
dk.altavista	9	0,02%

Referring search engines in KID (top 10).

Search Engine	#	%
google	11802	96,47%
webcache.googleusercontent	104	0,85%
bing	88	0,72%
translate.googleusercontent	83	0,68%
search.yahoo	67	0,55%
find.tdc	22	0,18%
dk.search.yahoo	14	0,11%
uk.search.yahoo	5	0,04%
se.search.yahoo	3	0,02%
dk.yhs.search.yahoo	3	0,02%
toolbar.google	3	0,02%

Referring search engines in Poma (top 10).

Search Engine	#	%
google	19769	95,13%
bing	298	1,43%
images.google	279	1,34%
translate.googleusercontent	152	0,73%
search.yahoo	70	0,34%
images.search.yahoo	53	0,26%
translate.google	41	0,20%
webcache.googleusercontent	30	0,14%
es.search.yahoo	15	0,07%
mx.search.yahoo	13	0,06%

Appendix 10 Object attributes

Evaluation results of object attributes in ADL.

Object			Number of SAPs			Full text (y/n) y=1 point	Comments
Id	Level	Title	None (0 SAPs) 0 points	Few (1-10 SAPs) 1 point	Many (11+ SAPs) 2 points		
AN1	N	Top page (first)		1			The only text: “Arkiv for Dansk Litteratur”
A-N2	N	Author list			2		All author names
A-N3	N	Search, page one, introduction			2		
A-N4	N	Search, page two, free text in texts and author biographies			2		
A-I1	I	Author first page: H.C. Andersen			2		
A-I2	I	Author title list: H.C. Andersen: D			2		
A-I3	I	Versions of the Little Mermaid			2		
A-I4	I	Author first page: Jacob Worm			2		
A-I5	I	Author title list: Jacob Worm			2		
A-I6	I	Versions of Annike Bi		1			
A-O1	O	The Little Mermaid, facimile, p.1.		1			Picture of text
A-O2	O	The Little Mermaid, facimile, p.3.		1			Picture of text
A-O3	O	The Little Mermaid, text, p.1.			2	1	
A-O4	O	The Little Mermaid, text, p.3.			2	1	
A-O5	O	The Little Mermaid, downloadable text			1	1	
A-O6	O	Annikke Bi, facimile		1			Picture of text

Evaluation results of object attributes in KID.

Object			Number of SAPs			Full text (y/n) y=1 points	Comments
Id	Level	Title	None (0 SAPs) 0 points	Few (1-10 SAPs) 1 point	Many (11+ SAPs) 2 points		
K-N1	N	Top page			2		
K-N2	N	Advanced search for artists			2		
K-N3	N	About Kunstindeks Danmark			2		
K-N4	N	Museums in Kunstindeks Danmark [A]			2		
K-I1	I	Artist: Karen Abell			2		
K-I2	I	List of artwork by Karen Abell		1			Depends on the number of works
K-I3	I	Information from Weilbachs Kunstnerleksikon: Karen Abell (all) (categories: genealogy, exhibitions, travels, education, occupations, biography, artworks, scholarships, literature, all)			2	1	
K-I4	I	AROS – Aarhus Kunstmuseum		1			
K-I5	I	Artist: F.M.E. Fabritius De Tengangel			2		
K-I6	I	Information from Weilbachs Kunstnerleksikon: F.M.E. Fabritius De Tengangel (all) (categories: genealogy, exhibitions, travels, education, occupations, biography, artworks, scholarships, literature, all)			2	1	
K-I7	I	Artist: Berenice Abbott		1			
K-O1	O	Artwork 1 by Karen Abell (“Uden titel”)			2		
K-O2	O	Artwork 4 by Geskel Saloman (“Portræt af Moses og Hanne Ruben“)			2		+ picture
K-O3	O	Artwork 14 by Fabritius de Tengnagel, F.M.E. (”Vinterlandskab fra Langeland”)			2		
K-O4	O	André Maurois by Berenice Abbott			2		+ picture

Evaluation results of object attributes in Poma.

Object			Number of SAPs			Full text (y/n) y=1 points	Comments
Id	Level	Title	None (0 SAPs) 0 points	Few (1-10 SAPs) 1 point	Many (11+ SAPs) 2 points		
P-N1	N	Top page (frontpage) English			2		
P-N2	N	Top page (frontpage) Spanish			2		
P-N3	N	Digital resources - English			2		
P-N4	N	Digital resources – Spanish (Recursos digitales)			2		
P-O1	O	Title page (Drawing 0 [1]) - English			2	1	
P-O2	O	Title page (Drawing 0 [1]) - Spanish			2	1	
P-O3	O	Page 1 - English			2	1	
P-O4	O	Page 1 - Spanish			2	1	
P-O5	O	Page 2 (Drawing 2) - English			2		
P-O6	O	Page 2 (Drawing 2) - Spanish			2		
P-O7	O	Page 79 (Drawing 23) (first page of Chapter 6) - English			2		
P-O8	O	Page 79 (Drawing 23) (first page of Chapter 6) - Spanish			2		
P-O9	O	Page 80 - English			2	1	
P-O10	O	Page 80 - Spanish			2	1	

Appendix 11 Accessibility

The accessibility measurement is a test of compliance to the WCAG 2.0 (level AA). The test is done by submitting the URL to an online Web Accessibilty checker called AChecker (http://achecker.ca). The numbers in the tables below is the output in the form of: number of known problems / number of likely problems / number of potential problems.

Evaluation results of accessibility in ADL.

Object			Number of WCAG-errors			
Id	Level	Title	To many (no compliance) 0 points	Many 1 point	Few 2 points	None (full compliance) 3 points
A-N1	N	Top page (first)			3/3/52	
A-N2	N	Author list			4/7/315	
A-N3	N	Search, page one, introduction			16/7/80	
A-N4	N	Search, page two, free text in texts and author biographies			33/8/106	
A-I1	I	Author first page: H.C. Andersen			29/7/94	
A-I2	I	Author title list: H.C. Andersen: D			7/7/315	
A-I3	I	Versions of the Little Mermaid			7/7/105	
A-I4	I	Author first page: Jacob Worm			20/7/88	
A-I5	I	Author title list: Jacob Worm			7/7/117	
A-I6	I	Versions of Annike Bi			7/7/110	
A-O1	O	The Little Mermaid, facimile, p.1.			8/11/95	
A-O2	O	The Little Mermaid, facimile, p.3.			14/0/18	
A-O3	O	The Little Mermaid, text, p.1.			8/11/95	
A-O4	O	The Little Mermaid, text, p.3.			17/0/20	
A-O5	O	The Little Mermaid, downloadable text				0
A-O6	O	Annike Bi, facimile			14/0/18	

Evaluation results of accessibility in KID.

Object			Number of WCAG-errors			
Id	Level	Title	To many (no compliance) 0 points	Many 1 point	Few 2 points	None (full compliance) 3 points
K-N1	N	Top page			13/0/135	
K-N2	N	Advanced search for artists			14/0/171	
K-N3	N	About Kunstindeks Danmark			1/0/710	
K-N4	N	Museums in Kunstindeks Danmark [A]			2/1/315	
K-I1	I	Artist: Karen Abell			9/0/178	
K-I2	I	List of artwork by Karen Abell			4/1/123	
K-I3	I	Information from Weilbachs Kunstnerleksikon: Karen Abell (all) (categories: genealogy, exhibitions, travels, education, occupations, biography, artworks, scholarships, literature, all)			1/0/96	
K-I4	I	AROS – Aarhus Kunstmuseum			5/0/90	
K-I5	I	Artist: F.M.E. Fabritius De Tengangel			9/0/178	
K-I6	I	Information from Weilbachs Kunstnerleksikon: F.M.E. Fabritius De Tengangel (all) (categories: genealogy, exhibitions, travels, education, occupations, biography, artworks, scholarships, literature, all)			1/0/96	
K-I7	I	Artist: Berenice Abbott			8/0/87	
K-O1	O	Artwork 1 by Karen Abell (“Uden titel”)			10/0/106	
K-O2	O	Artwork 4 by Geskel Saloman (“Portræt af Moses og Hanne Ruben“)			11/0/149	
K-O3	O	Artwork 14 by Fabritius de Tengnagel, F.M.E. (”Vinterlandskab fra Langeland”)			9/0/106	
K-O4	O	André Maurois by Berenice Abbott			8/0/115	

Evaluation results of accessibility in Poma.

Object			Number of WCAG-errors			
Id	Level	Title	To many (no compliance) 0 points	Many 1 point	Few 2 points	None (full compliance) 3 points
P-N1	N	Top page (frontpage) English			62/1/970	
P-N2	N	Top page (frontpage) Spanish			62/1/967	
P-N3	N	Digital resources - English			6/1/860	
P-N4	N	Digital resources – Spanish (Recursos digitales)			6/1/967	
P-O1	O	Title page (Drawing 0 [1]) - English			6/1/860	
P-O2	O	Title page (Drawing 0 [1]) - Spanish			6/1/857	
P-O3	O	Page 1 - English			6/1/872	
P-O4	O	Page 1 - Spanish			6/1/869	
P-O5	O	Page 2 (Drawing 2) - English			6/1/857	
P-O6	O	Page 2 (Drawing 2) - Spanish			6/1/854	
P-O7	O	Page 79 (Drawing 23) (first page of Chapter 6) - English			6/1/857	
P-O8	O	Page 79 (Drawing 23) (first page of Chapter 6) - Spanish			6/1/852	
P-O9	O	Page 80 - English			6/1/862	
P-O10	O	Page 80 - Spanish			6/1/859	

Appendix 12 Internal navigation and internal search

Evaluation results of internal navigation and internal search in ADL.

Object			Possible to follow links to object (y/n) y=1 point	Possible to find object through internal search engine (y/n) y=1 point
Id	Level	Title		
A-N1	N	Top page (first)	1	0
A-N2	N	Author list	1	0
A-N3	N	Search, page one, introduction	1	0
A-N4	N	Search, page two, free text in texts and author biographies	1	0
A-I1	I	Author first page: H.C. Andersen	1	1
A-I2	I	Author title list: H.C. Andersen: D	1	1 (to A, then click on D)
A-I3	I	Versions of the Little Mermaid	1	1 (in Title search)
A-I4	I	Author first page: Jacob Worm	1	1
A-I5	I	Author title list: Jacob Worm	1	1
A-I6	I	Versions of Annike Bi	1	1 (in Title search)
A-O1	O	The Little Mermaid, facimile, p.1.	1	0 (not in Text-page search)
A-O2	O	The Little Mermaid, facimile, p.3.	1	0 (not in Text-page search)
A-O3	O	The Little Mermaid, text, p.1.	1	1
A-O4	O	The Little Mermaid, text, p.3.	1	1
A-O5	O	The Little Mermaid, downloadable text	1	0 (not in Text-page search)
A-O6	O	Annike Bi, facimile	1	0 (not in Text-page search)

Evaluation results of internal navigation and internal search in KID.

Object			Possible to follow links to object (y/n) y=1 point	Possible to find object through internal search engine (y/n) y=1 point
Id	Level	Title		
K-N1	N	Top page	1	0
K-N2	N	Advanced search for artists	1	0
K-N3	N	About Kunstindeks Danmark	1	0
K-N4	N	Museums in Kunstindeks Danmark [A]	1	0
K-I1	I	Artist: Karen Abell	1	1
K-I2	I	List of artwork by Karen Abell	1	1
K-I3	I	Information from Weilbachs Kunstnerleksikon: Karen Abell (all) (categories: genealogy, exhibitions, travels, education, occupations, biography, artworks, scholarships, literature, all)	1	0 (not direct, through "search for artist")
K-I4	I	AROS – Aarhus Kunstmuseum	1	0
K-I5	I	Artist: F.M.E. Fabritius De Tengangel	1	1
K-I6	I	Information from Weilbachs Kunstnerleksikon: F.M.E. Fabritius De Tengangel (all) (categories: genealogy, exhibitions, travels, education, occupations, biography, artworks, scholarships, literature, all)	1	0 (not direct, through "search for artist")
K-I7	I	Artist: Berenice Abbott	1	1
K-O1	O	Artwork 1 by Karen Abell ("Uden titel")	1	1
K-O2	O	Artwork 4 by Geskel Saloman ("Portræt af Moses og Hanne Ruben")	1	1
K-O3	O	Artwork 14 by Fabritius de Tengnagel, F.M.E. ("Vinterlandskab fra Langeland")	1	1
K-O4	O	André Maurois by Berenice Abbott	1	1

Evaluation results of internal navigation and internal search in Poma.

Object			Possible to follow links to object (y/n) y=1 point	Possible to find object through internal search engine (y/n) y=1 point
Id	Level	Title		
P-N1	N	Top page (frontpage) English	1	0
P-N2	N	Top page (frontpage) Spanish	1	0
P-N3	N	Digital resources - English	1	0
P-N4	N	Digital resources – Spanish (Recursos digitales)	1	0
P-O1	O	Title page (Drawing 0 [1]) - English	1	1
P-O2	O	Title page (Drawing 0 [1]) - Spanish	1	1
P-O3	O	Page 1 - English	1	1
P-O4	O	Page 1 - Spanish	1	1
P-O5	O	Page 2 (Drawing 2) - English	1	1
P-O6	O	Page 2 (Drawing 2) - Spanish	1	1
P-O7	O	Page 79 (Drawing 23) (first page of Chapter 6) - English	1	1
P-O8	O	Page 79 (Drawing 23) (first page of Chapter 6) - Spanish	1	1
P-O9	O	Page 80 - English	1	1
P-O10	O	Page 80 - Spanish	1	1

Appendix 13 Reachability

Evaluation results of reachability in ADL.

Object			Possible to link to object (y/n) y=1 points
Id	Level	Title	
A-N1	N	Top page (first)	1
A-N2	N	Author list	1
A-N3	N	Search, page one, introduction	1
A-N4	N	Search, page two, free text in texts and author biographies	1
A-I1	I	Author first page: H.C. Andersen	1
A-I2	I	Author title list: H.C. Andersen: D	1
A-I3	I	Versions of the Little Mermaid	1
A-I4	I	Author first page: Jacob Worm	1
A-I5	I	Author title list: Jacob Worm	1
A-I6	I	Versions of Annike Bi	1
A-O1	O	The Little Mermaid, facimile, p.1.	1
A-O2	O	The Little Mermaid, facimile, p.3.	1
A-O3	O	The Little Mermaid, text, p.1.	1
A-O4	O	The Little Mermaid, text, p.3.	1
A-O5	O	The Little Mermaid, downloadable text	1
A-O6	O	Annike Bi, facimile	1

Evaluation results of reachability in KID.

Object			Possible to link to object (y/n) y=1 points
Id	Level	Title	
K-N1	N	Top page	1
K-N2	N	Advanced search for artists	1
K-N3	N	About Kunstindeks Danmark	1
K-N4	N	Museums in Kunstindeks Danmark [A]	1
K-I1	I	Artist: Karen Abell	1
K-I2	I	List of artwork by Karen Abell	1
K-I3	I	Information from Weilbachs Kunstnerleksikon: Karen Abell (all) (categories: genealogy, exhibitions, travels, education, occupations, biography, artworks, scholarships, literature, all)	1
K-I4	I	AROS – Aarhus Kunstmuseum	1
K-I5	I	Artist: F.M.E. Fabritius De Tengangel	1
K-I6	I	Information from Weilbachs Kunstnerleksikon: F.M.E. Fabritius De Tengangel (all) (categories: genealogy, exhibitions, travels, education, occupations, biography, artworks, scholarships, literature, all)	1
K-I7	I	Artist: Berenice Abbott	1
K-O1	O	Artwork 1 by Karen Abell (“Uden titel”)	1
K-O2	O	Artwork 4 by Geskel Saloman (“Portræt af Moses og Hanne Ruben“)	1
K-O3	O	Artwork 14 by Fabritius de Tengnagel, F.M.E. (”Vinterlandskab fra Langeland”)	1
K-O4	O	André Maurois by Berenice Abbott	1

Evaluation results of reachability in Poma.

Object			Possible to link to object (y/n) y=1 points
Id	Level	Title	
P-N1	N	Top page (frontpage) English	1
P-N2	N	Top page (frontpage) Spanish	1
P-N3	N	Digital resources - English	1
P-N4	N	Digital resources – Spanish (Recursos digitales)	1
P-O1	O	Title page (Drawing 0 [1]) - English	1
P-O2	O	Title page (Drawing 0 [1]) - Spanish	1
P-O3	O	Page 1 - English	1
P-O4	O	Page 1 - Spanish	1
P-O5	O	Page 2 (Drawing 2) - English	1
P-O6	O	Page 2 (Drawing 2) - Spanish	1
P-O7	O	Page 79 (Drawing 23) (first page of Chapter 6) - English	1
P-O8	O	Page 79 (Drawing 23) (first page of Chapter 6) - Spanish	1
P-O9	O	Page 80 - English	1
P-O10	O	Page 80 - Spanish	1

Appendix 14 Web prestige

Evaluation results of web prestige in ADL.

Object			PageRank-value				Points Web prestige
Id	Level	Title	No value (not indexed/-ranked) 0 points	Low (0-2) 1 points	Medium (3-5) 2 points	High (6-10) 3 points	
A-N1	N	Top page (first)				7	**3**
A-N2	N	Author list				6	**3**
A-N3	N	Search, page one, introduction				6	**3**
A-N4	N	Search, page two, free text in texts and author biographies				6	**3**
A-I1	I	Author first page: H.C. Andersen				7	**3**
A-I2	I	Author title list: H.C. Andersen: D			5		**2**
A-I3	I	Versions of the Little Mermaid	x				**0**
A-I4	I	Author first page: Jacob Worm			5		**2**
A-I5	I	Author title list: Jacob Worm	X, indexed and cashed, not ranked				**0**
A-I6	I	Versions of Annike Bi	X, not indexed, cashed or ranked				**0**
A-O1	O	The Little Mermaid, facimile, p.1.			5		**2**
A-O2	O	The Little Mermaid, facimile, p.3.	X, indexed and cashed, not ranked				**0**
A-O3	O	The Little Mermaid, text, p.1.			4		**2**
A-O4	O	The Little Mermaid, text, p.3.	X, indexed and cashed, not ranked				**0**
A-O5	O	The Little Mermaid, downloadable text	X, not indexed, cashed or ranked				**0**
A-O6	O	Annike Bi, facimile	X, not indexed, cashed or ranked				**0**

Evaluation results of web prestige in KID.

Object			PageRank-value				Points Web prestige
Id	Level	Title	No value (not indexed/ranked) [0 points]	Low (0-2) 1 points	Medium (3-5) 2 points	High (6-10) 3 points	
K-N1	N	Top page	X, Indexed and cashed, not ranked				**0**
K-N2	N	Advanced search for artists	X, Indexed and cashed, not ranked				**0**
K-N3	N	About Kunstindeks Danmark	X, Indexed and cashed, not ranked				**0**
K-N4	N	Museums in Kunstindeks Danmark [A]	X, Indexed and cashed, not ranked				**0**
K-I1	I	Artist: Karen Abell	X, Indexed and cashed, not ranked				**0**
K-I2	I	List of artwork by Karen Abell	X, Indexed and cashed, not ranked				**0**
K-I3	I	Information from Weilbachs Kunstnerleksikon: Karen Abell (all) (categories: genealogy, exhibitions, travels, education, occupations, biography, artworks, scholarships, literature, all)	X, Indexed and cashed, not ranked				**0**
K-I4	I	AROS – Aarhus Kunstmuseum	X, Indexed and cashed, not ranked				**0**
K-I5	I	Artist: F.M.E. Fabritius De Tengangel	X, Indexed and cashed, not ranked				**0**
K-I6	I	Information from Weilbachs Kunstnerleksikon: F.M.E. Fabritius De Tengangel (all) (categories: genealogy, exhibitions, travels, education, occupations, biography, artworks, scholarships, literature, all)	X, Indexed and cashed, not ranked				**0**
K-I7	I	Artist: Berenice Abbott	X, Indexed and cashed, not ranked				**0**
K-O1	O	Artwork 1 by Karen Abell (“Uden titel”)	X, Indexed and cashed, not ranked				**0**
K-O2	O	Artwork 4 by Geskel Saloman (“Portræt af Moses og Hanne Ruben“)	X, Indexed and cashed, not ranked				**0**
K-O3	O	Artwork 14 by Fabritius de Tengnagel, F.M.E. (”Vinterlandskab fra Langeland”)	X, Indexed and cashed, not ranked				**0**
K-O4	O	André Maurois by Berenice Abbott	X, Indexed and cashed, not ranked				**0**

Evaluation results of web prestige in Poma.

Object			PageRank-value				Points Web prestige
Id	Level	Title	No value (not indexed/-ranked) 0 points	Low (0-2) 1 points	Medium (3-5) 2 points	High (6-10) 3 points	
P-N1	N	Top page (frontpage) English			5		**2**
P-N2	N	Top page (frontpage) Spanish				6	**3**
P-N3	N	Digital resources - English			3		**2**
P-N4	N	Digital resources – Spanish (Recursos digitales)			4		**2**
P-O1	O	Title page (Drawing 0 [1]) - English	Indexed and cashed, not ranked				**0**
P-O2	O	Title page (Drawing 0 [1]) - Spanish			4		**2**
P-O3	O	Page 1 - English			4		**2**
P-O4	O	Page 1 - Spanish			4		**2**
P-O5	O	Page 2 (Drawing 2) - English	X, Indexed and cashed, not ranked				**0**
P-O6	O	Page 2 (Drawing 2) - Spanish	X, Indexed and cashed, not ranked				**0**
P-O7	O	Page 79 (Drawing 23) (first page of Chapter 6) - English			3		**2**
P-O8	O	Page 79 (Drawing 23) (first page of Chapter 6) - Spanish	X, Indexed and cashed, not ranked				**0**
P-O9	O	Page 80 - English	X, Indexed and cashed, not ranked				**0**
P-O10	O	Page 80 - Spanish	X, Indexed and cashed, not ranked				**0**

Appendix 15 Navigation strategies in logs

Navigation strategies for the period October 20102 to December 2010.

	ADL Log (n=44352)	ADL %	KID Log (n=22667)	KID %	Poma Log (n=30557)	Poma %
Direct	3973	9.0%	3750	17%	1725	5.6%
Link	2514	5.7%	6941	31%	8974	29.4%
Search Engine	37865	85.4%	11976	53%	19858	65.0%
Total	*44352*	*100%*	*22667*	*100%*	*30557*	*100%*

Appendix 16 Queries in referring search engines

The first three tables lists the most frequent queries in the resources.

The last three tables lists the queries in the samples from the whole distributions.

ADL referring search engine phrases (top 20).

Phrase	#	Type
adl	302	N
herman bang	302	I
ludvig holberg	282	I
arkiv for dansk litteratur	277	N
leonora christine	213	I
emil aarestrup	197	I
dansk litteratur	174	I (N)
søren kierkegaard	145	I
danske forfattere	135	I
thomas kingo	131	I
adl.dk	127	N
steen steensen blicher	125	I
irene holm	123	I
sexnoveller	121	I
herman bang frøkenen	119	I
erasmus montanus	116	I
herman bang pernille	114	I
jeppe aakjær	111	I
den grimme Ælling tekst	107	I
georg brandes	88	I

KID referring search engine phrases (top 20).

Phrase	#	Type
weilbachs kunstnerleksikon	571	N
kid	344	N
weilbach	170	N
kid.dk	115	N
kunstindeks danmark	107	N
www.kid.dk	67	N
lars swane	53	I
weilbachs	51	N
kunstnerleksikon	40	N (I)
kunstindex danmark	37	N
weilbach kunstnerleksikon	33	N
sigurd swane	33	I
kunstleksikon	28	I (N)
johannes carstensen	27	I
gudmund olsen	26	I
hans brygge	26	I
weilbach kunst	26	N
johannes larsen	24	I
danske billedkunstnere	24	I
lars svane	21	I

Poma referring search engine phrases (top 20).

Phrase	#	Type
guaman poma	670	N (I)
guaman poma de ayala	507	I (N)
felipe guaman poma de ayala	165	I (N)
alcaldes mayores	143	I
nueva coronica y buen gobierno	129	I (N)
muerte de pizarro	109	I
vestimenta de los españoles	104	I
vestimenta de los españoles en la conquista	65	I
cronicas de guaman poma de ayala	61	I (N)
huaman poma de ayala	53	I
el sitio de guaman poma	52	N
guaman poma de ayala dibujos	51	I
poma de ayala	47	I (N)
guaman poma website	43	N
guaman poma de ayala cronicas	42	I (N)
segundo virrey del peru	42	I
biblioteca real de copenhague	41	N
milagros de dios	40	I
nueva cronica y buen gobierno	39	I (N)
vestimenta de los conquistadores españoles	36	I

ADL referring search engine phrases (sample of every hundredth).

Query	**#**	TRANS	NAV	INFO	I-CR	I-SO	I-FU	I-GEO	I-GEN	I-TP	I-SI	I-TO	I-TY	I-OT	Poly
adl	302		N												
kingo	29			I	CR										
henrik hertz sparekassen	15			I	CR	SO									X
ernesto dalgas	11			I	CR										
senklassicisme	8			I						TP					
herman bang frøkenen analyse	7			I	CR	SO			GEN						X
sundt blod	6			I		SO									
hjorterytteren	5			I		SO									
ludvig bødtcher	4			I	CR										
den gamle præst	4			I		SO									
tragedie af oehlenschlager	4			I	CR				GEN						X
fortolkning og analyse af a.w. schack von staffeldt digt til naturen	3			I	CR	SO			GEN						X
gaffelen	3			I		SO									
her har jeg himlene over min isse	3			I			FU								
herman bang chopin	3			I	CR	SO									X
reflexion emil aarestrup	2			I	CR				GEN						X
dødens gudsøn	2			I								TO			
epistola cclvii	2			I		SO									
holberg erasmus montanus	2			I	CR	SO									X
paris skytshelgen	2			I			FU								
dåbsdigt	2			I					GEN			TO			X
kærlighed af herman bang	2			I	CR	SO									X
poe og romantismen	2			I	CR					TP					X
jammers minde analyse	2			I		SO			GEN						X
præludium steen steensen blicher analyse	2			I	CR	SO			GEN						X
en middag emil aarestrup	2			I	CR	SO									X
teksten frøkenen	2			I		SO							TY		X
rokoko musik	2			I						TP			TY		X
kat skorpionbid	1			I			FU								
herman bang novelle adl	1		N	I	CR				GEN						X
en god samvittighed komposition	1			I										OT	
hvem er pernille af herman bang	1			I	CR	SO									X
klassicismes komedie	1			I					GEN	TP					X
1 kor.4.1-5	1			I										OT	
ufødt fremtid ned fra himlen stiger	1			I			FU								

Query	**#**	TRANS	NAV	INFO	I-CR	I-SO	I-FU	I-GEO	I-GEN	I-TP	I-SI	I-TO	I-TY	I-OT	Poly
børne sangersn	1			I								TO			
kirkegård digt	1			I	CR				GEN						X
herman bang frøkenen	1			I	CR	SO									X
sangen fra svanen og trompeten vi flyver bort	1			I			FU		GEN						X
modtagelse samtid erasmus montanus	1			I		SO								OT	X
1870-90 "det moderne gennembrud"	1			I						TP		TO			X
emil aarestrup uddrag	1			I	CR	SO									X
bibliografi hvad skal i kursiv?	1			I					GEN					OT	X
adam oehlenschläger sex	1			I	CR							TO			X
problem om arvesynden	1			I								TO			
ludvig holberg epistel 46	1			I	CR	SO									X
hovedgaard sigersted	1			I								TO			
hc andersen stil	1			I	CR									OT	X
tycho brahe søn af	1			I	CR							TO			X
h.c. andersen dansker	1			I	CR	SO									X
antal danske digtere siden 1600 tallet	1			I				GEO		TP				OT	X
café styrker maven, hindrer dunsterne at opstige, og conseqventer stiller saa vel tand- som hoved-pine, hvilket jeg saa ofte haver erfaret, at jeg gandske er bleven overbeviset derom.	1			I			FU								
otto borchsenius drachmann	1			I	CR							TO			X
heiberg lyrik	1			I	CR				GEN						X
billeder af lys på ludvig	1			I	CR							TO			X
livets mening, af gustaf munch petersen	1			I	CR	SO									X
carl bagger roman digt	1			I	CR				GEN						X
nævner min tanke dig	1			I		SO									
julesalmer thomas kingo	1			I	CR				GEN						X
grundtvig kendte værker	1			I	CR									OT	X
bryllupsnatten, gammel dansk digt	1			I		SO			GEN					OT	X
et skud i tågen lydbånd	1			I		SO							TY		X
naturalisme lykke-per	1			I		SO			GEN						X
nattergalen h c andersen	1			I	CR	SO									X
digt - ingeborg	1			I					GEN			TO			X
et brev om darwinismen	1			I		SO									
kirkegård filosof	1			I	CR									OT	X
dannemarks og norges beskrivelse holberg	1			I	CR			GEO	GEN						X

Query	**#**	TRANS	NAV	INFO	I-CR	I-SO	I-FU	I-GEO	I-GEN	I-TP	I-SI	I-TO	I-TY	I-OT	Poly
odins ledsager	1			I								TO			
spisekammer historie	1			I			FU								
christiane f barndom	1			I								TO			
grundtvig dagbøger	1			I	CR				GEN						X
st. st. blicher hosekræmmeren. panel drøftelse	1			I	CR	SO			GEN						X
jo større jo bedre	1			I			FU								
hestetrukken herregårds kane	1			I			FU								
jensen studier over h. c. andersens sprog	1			I	CR							TO			X
jørgen vosmar	1			I	CR										
nattens dæmrende tåger	1			I		SO									
kingo sorrig og glæde analyse den danske salmebog	1			I	CR	SO			GEN			TO			X
tvetydighed tante tandpine	1			I		SO						TO			X
de fire vinde aakjær	1			I	CR	SO									X
teksten til " digterjul "	1			I		SO							TY		X
jørgen hattemager salomon holberg	1			I	CR	SO									X
strik damehue fra ugebladet hjemmet	1			I								TO			
julebrev - kære elskede	1			I								TO			
en coopersk roman	1			I		SO									
du kom, og tog mig med storm, viste mig kærlighed i enhver form. fejede benene væk under mig, viste mig loyalitetens vej. og vi så ærligheden overtage, for så at bortviske uforglemmelige dage. alt er aldrig som man tror, og sandheden sætter altid brændende spor. desværre..	1			I			FU								
forfatterweb erik skram	1		N	I	CR										
vandklar kirsebærbrændevin	1			I			FU								
små børnevers	1			I					GEN					OT	X
hvilke forhold i samfundet danner grundlag for naturalismen	1			I								TO			
danskeren iv	1			I		SO									
huggo.dk	1		N												
sibirisk mår	1			I								TO			
klokkestreng tjener	1			I			FU								
jens lassen knudsen	1			I	CR										
thøger larsen du danske sommer	1			I	CR		FU								X

Query	**#**	TRANS	NAV	INFO	I-CR	I-SO	I-FU	I-GEO	I-GEN	I-TP	I-SI	I-TO	I-TY	I-OT	Poly
romantikken som periode i dansk litteratur	1			I						TP		TO			X
analyse af udløbet i uendelighed	1			I		SO			GEN						X
hofprædikant bluhme frederik d. 4.	1			I								TO			
møller-christensen, ivy york (1992): den gyldne trekant. h.c. andersens gennembrud i tyskland 1831-1850	1			I								TO			
nattens dæmrende tåge	1			I		SO									
oplysningsmenneske	1			I								TO			
harry jacobsen	1			I								TO			
violera_26	1			I			FU								
efterstykke holberg	1			I	CR	SO									X
lyden af dig selv af christian winther	1			I	CR	SO									X
b.s. ingemann periode	1			I	CR									OT	X
stearin på plyssofa	1			I			FU								
saxo grammaticus 1818	1			I	CR	SO				TP					X
henrik pontoppidan - realist	1			I	CR									OT	X
rationalisme og pietisme	1			I						TP		TO			X
biedermeier christian træsnit	1			I	CR	SO						TO			X
skolesang	1			I		SO									
h. c. andersen klokken ironi	1			I	CR	SO									X
jp to verdner	1			I	CR	SO									X
kendte danske skønlitterær tekster h.c. andersen	1			I	CR							TO			X
hans christian andersens forfatterskab	1			I	CR							TO			X
to verdener (en skitse)	1			I		SO									
kingo anden part	1			I	CR	SO									X
en jøde 1845	1			I		SO				TP					X
man har sagn om	1			I			FU								
søren kanne buhl	1			I	CR										
saxos danmarks historie	1			I	CR	SO									X
gustav hvid hofjægermester	1			I	CR	SO									X
smerter i skulderen og netsat bevægelighed	1			I			FU								
imerco ønskeseddel dina og jens	1			I										OT	
det døende barn hc andersen	1			I	CR	SO									X
oehlenschlÃ¤ger elskovskunst	1			I	CR	SO									X
jp jacobsen marine	1			I	CR	SO									X
vielsestale	1			I		SO									
adam øens	1			I	CR										

Query	**#**	TRANS	NAV	INFO	I-CR	I-SO	I-FU	I-GEO	I-GEN	I-TP	I-SI	I-TO	I-TY	I-OT	Poly
tragedie af adam oehlenschlæger	1			I	CR				GEN						X
johannes weltzer	1			I	CR										
ambrosius stub naturopfattelse	1			I	CR							TO			X
recipere	1			I										OT	
litteratur romantik	1			I					GEN	TP					X
vigtige forfattere i 1956-70	1			I						TP		TO			X
hostrups komedier	1			I	CR				GEN						X
bernhard severin ingemann familie	1			I	CR							TO			X
fru fønss fortolkning	1			I		SO			GEN						X
den sidste dag i det 18. århundrede	1			I			FU								
reaktionen	1			I		SO									
skyggen af st. st. blicher	1			I	CR	SO									X
digte om skilsmisser	1			I								TO			
biografi kaj munk aakjær andersen	1			I					GEN			TO			X
bærorm	1			I										OT	
peter willemoes avis udkast	1			I										OT	
niels klim analyse	1			I	CR				GEN						X
cv dag encke	1			I										OT	
den hellig ånd gustav wied	1			I	CR	SO									X
tekst den lille havfrue	1			I		SO							TY		X
sophus claussen kærlighed	1			I	CR	SO									X
christian winther en aftenscene	1			I	CR	SO									X
steen steensen blicher himmelbjerg	1			I	CR	SO									X
talemåde-j.c. christensens valgsprog	1			I	CR							TO			
gustav munch petersen mod jerusalem	1			I	CR	SO									X
den suhrske familie	1			I		SO									
frøding	1			I	CR										
jeppe aakjær naar rugen skal ind	1			I	CR	SO									X
ave maria ydmygt knæler stille besjæler	1			I			FU								
kingos "hosianna"	1			I	CR	SO									X
ludvig holberg fabel gris	1			I	CR	SO									X
de dumme dänen slogan	1			I			FU		GEN						X

Query	**#**	TRANS	NAV	INFO	I-CR	I-SO	I-FU	I-GEO	I-GEN	I-TP	I-SI	I-TO	I-TY	I-OT	Poly
mange kannibaler spiste noget af hjernen på deres bytte for at erhverve sig deres visdom, myter og vrede. i denne bog får du lidt af min hjerne at tygge på - og så benytter jeg mig af lejligheden til også at tage en bid af din	1			I			FU								
om at fortælle børn eventyr poul martin møller	1			I	CR	SO									X
litteratur perlen	1			I		SO							TY		X
digter døde i sorø	1			I		SO			GEN						X
arendse johannes ewald	1			I	CR										
digt aladdin lavet af adam	1			I	CR				GEN			TO			X
laps gammelt kortspil	1			I			FU								
herman bang journalistiske karriere	1			I								TO			
den tizianske magdalene	1			I								TO			
librum	1			I		SO									
kunstnerlængsel af schack von staffeldt analyse	1			I	CR	SO			GEN						X
moderne historie om den lille pige med svovlstikkerne	1			I								TO			
st.st. blicher født	1			I	CR							TO			X
1842 "han er mig kær"	1			I			FU								
digt om væv og skytte og livet	1			I					GEN			TO			X
danmark dejligst vang og vænge stakkels lille land,	1			I			FU								
oehlenschläger "karl den store"	1			I	CR	SO									X
christian 4 leonora	1			I	CR									OT	X
dansk børnesang gubben noah holdt så meget af den sov med den om natten	1			I				GEO	GEN			TO			X
litteratur industrialisering konsekvenser	1			I								TO	TY		X
analyse af candida af kingo	1			I	CR	SO			GEN						X
h.c. andersen romantisme værker	1			I	CR				GEN	TP					X
korte børnedigte	1			I					GEN					OT	X
grossererballer	1			I								TO			
konge af sverige i 1600	1			I				GEO		TP		TO			X
canova erotisk	1			I								TO			
olden-jørgensen helgesen	1			I	CR										

KID referring search engine phrases (sample of every hundredth).

Query	#	TRANS	NAV	INFO	I-CR	I-SO	I-FU	I-GEO	I-GEN	I-TP	I-SI	I-TO	I-TY	I-OT	Poly
weilbachs kunstnerleksikon	571		N												
frants landt	6			I	CR										
gunnar hossy	4			I	CR										
viggo kragh-hansen	3			I	CR										
ulvig	2			I	CR										
optagelse i weilbach	2			I										OT	
arkitekt arne arcel	2			I	CR									OT	X
giersing selvportræt	2			I	CR				GEN						X
erik ejlers	2			I	CR										
kunst malere fra fyn 1889	1			I				GEO	GEN	TP					X
erik damgaard henriksen, købes	1			I	CR										
christen dalsgaard mon han dog ikke skulle komme	1			I	CR	SO									X
laurits tuxen 1895 skt hans	1			I	CR					TP		TO			X
kid, dk	1		N												
dansk kunst skib på havet bille	1			I				GEO				TO			X
selvportræt med cigar jens sørensen	1			I	CR				GEN					OT	X
aften og erotik fritz syberg	1			I	CR	SO									X
skagensmalere 1901	1			I					GEN	TP					X
skovgaardmuseet og bendt thoft nielsen	1			I	CR						SI				X
knud hendriksen xylograf	1			I	CR				GEN						X
backhausen maler	1			I	CR				GEN						X
stedelijk museum, amsterdam 1949	1			I				GEO		TP	SI				X
christian borg junker keramik	1			I	CR								TY		X
fru n p bolt	1			I	CR	SO									X
ernst eberlein billedhugger	1			I	CR				GEN						X
otto holm	1			I	CR										
alfred simonsen	1			I	CR										
frits bruzelius	1			I	CR										
carl madsen maler	1			I	CR									OT	X
tilstandstryk	1			I										OT	
udslidt 1889 af h. a. brendekilde	1			I	CR	SO				TP					X
marinemaler jens erik carl rasmussen	1			I	CR				GEN						X
bøgeskov i maj. motiv fra iselingen,	1			I		SO									
p s krøyer 1898	1			I	CR					TP					X
andreas magerstadt painter denmark	1			I	CR			GEO	GEN						X

Query	#	TRANS	NAV	INFO	I-CR	I-SO	I-FU	I-GEO	I-GEN	I-TP	I-SI	I-TO	I-TY	I-OT	Poly
gamle pax kerteminde	1			I										OT	
willumsen grafik	1			I	CR								TY		X
dødsfald fritz syberg	1			I	CR	SO									X
n. larsens stevns, kristus	1			I	CR							TO			X
johannes evangelisten thorvaldsen	1			I	CR	SO									X
stregfigur	1			I								TO			
maleren hans dall	1			I	CR				GEN						X
odense skulptur græske pige med stor krukke	1			I				GEO				TO	TY		X
tolkning af wilhelm freddie kongernes konge	1			I	CR	SO			GEN						X
christine østergaard	1			I	CR										
paludan madsen konst marine	1			I	CR							TO			X
udsigt fra dosseringen ved sortedamssøen mod nørrebro, maleri af købke (1838)	1			I	CR	SO				TP			TY		X
per kirkeby 1974	1			I	CR					TP					X
søholm klampenborg arne jacobsen	1			I	CR			GEO							X
kukkenbjergvejen	1			I		SO									
billedhugger yan	1			I	CR							TO			X
fritz stær olsen	1			I	CR										
alex steen	1			I	CR										
malerier malet af j.von holst	1			I	CR				GEN						X
parti fra iselingen skov, 1861	1			I		SO				TP					X
fritz syberg baron	1			I	CR									OT	X
lystighed udenfor roms mure på en oktober nat marstrand	1			I	CR	SO									X

Poma referring search engine phrases (sample of every hundredth).

Query	#	TRANS	NAV	INFO	I-CR	I-SO	I-FU	I-GEO	I-GEN	I-TP	I-SI-	I-TO	I-TY	I-OT	Poly
guaman poma	670		N	(I)											
la nueva coronica y buen gobierno	10		(N)	I		SO									
antonio de mendoza virrey del peru	6			I				GEO				TO		OT	X
fernando torres de portugal	4			I				GEO				TO			X
old women	3			I								TO			
guzmán poma de ayala imágenes	2			I								TO	TY		X
awakuq warmi	2			I			(FU)					TO			
dibujo mapa mundial	2			I								TO	TY		X
guaman poma el primer nueva coronica	2			I		SO									
conde villar arica	1			I								TO		OT	X
traje cristobal colon	1			I								TO			
mapa descubrimiento mar del sur	1			I								TO	TY		X
indumentaria masculina de los españoles	1			I								TO			
la feligresa tiene una penitencia con un cura capitulo 2	1			I			FU							OT	X
inca quiso yupanki	1			I			FU								
virrey fernando torres de portugal	1			I				GEO				TO		OT	X
auqui tupac	1			I								TO			
corregidor de provincias	1			I								TO			
200 year old men	1			I			FU								
don francisco of toledo	1			I				GEO				TO			X
postillon	1			I								TO			
guaman pom de ayala	1			I		SO									
isaac almanza amaro	1			I								TO			
supersticiones andinas la pulga	1			I								TO			
paria caca	1			I								TO			
nuev coronica y buen gobierno	1		(N)	I		SO									
vestimenta de los españoles	1			I				GEO				TO			X
el sitiod guaman poma	1		N												
felipe huamán poma de ayala	1			I								TO			
4 edades de la cronica de guaman poma	1			I		SO	FU								X
juan diaz de solis : sus vajes marcados en un mapa	1			I			FU					TO			X
guaman poma stio	1		N												
fotos del sacramento del matrimonio	1			I								TO	TY		X

Query	#	TRANS	NAV	INFO	I-CR	I-SO	I-FU	I-GEO	I-GEN	I-TP	I-SI-	I-TO	I-TY	I-OT	Poly
imagen sacramento del bautismo	1			I								TO	TY		X
vastimenta de francisco pizarro	1			I								TO			
imagenes de dios nuestro señor	1			I								TO	TY		X
poma de ayala los guerreros incas la primera calle o grupo de edad de hombres	1			I								TO			
los alcalde mayores	1			I								TO			
que es una mala confesion	1			I								TO			
guaman poma y la ilustración	1			I								TO	TY		X
que los dichos caciques principales y sus yndios o las yndias, sus propios hijos lexÃ-timos que dansen y hagan taquies y haylli [cantos] a: uacon uauco, saynata, llama llama, haya chuco, chimo capac, ayanya, guarmi auca, anti suyo, chipchi llanto, uaruro, hahiua, apac, llamaya, hara uayo, uaricza, tumi pampa, haraui, pingollo, quena quena, cata uari y dansas de espaÃ±oles y de negros y otras dansas de los yndios...	1			I								TO			
ordenanza 1-95 y 1-96	1			I								TO			
drawing atahualpa	1			I								TO	TY		X
segundo virreinato	1			I								TO			
fallecimiento de francisco toledo	1			I								TO			
http://www.kb.dk/elib/mss/poma fiesta de taki en quechua	1		N	I											
la nueva cronica del buen gobierno	1		(N)	I		SO									
milagros q hizo	1			I								TO			
poma de ayala año	1			I								TO			

Appendix 17 Referring sites (links group per site)

ADL referring links grouped on site (top 20).

Wikipedia page	Number of session
andersen.sdu	1081
da.wikipedia	990
Kb	540
Skablet	533
en.wikipedia	449
Bjoerna	325
Dsl	288
www2.kb	286
search.conduit	257
Facebook	239
Emu	236
Fronter	231
no.wikipedia	180
theeuropeanlibrary	142
seniorinternet	129
Vufintern	93
de.wikipedia	90
Eniro	86
Runeberg	83
web-reolen	79

KID referring links grouped on site (top 20).

Site	Number of session
Kunstonline	1571
da.wikipedia	650
Kunstmaler	564
signaturbogen.wikidot	394
Guldalder	339
en.wikipedia	247
Kunsten	230
skagensmalerne	151
skagensmuseum	149
Kb	109
no.wikipedia	88
Tisvildekunst	83
bornholms-kunstmuseum	82
search.conduit	69
Fyn	62
de.wikipedia	58
Kunstportalfyn	56
Hirschsprung	53
Kunstbib	52
galerie-ab	52

Poma referring links grouped on site (top 20).

Site	Number of session
abp-sil-colonia.blogspot	13841
cybertlink2.blogspot	3221
vidavirrey.blogspot	1669
search.conduit	373
es.wikipedia	370
en.wikipedia	340
chnm.gmu.edu	156
facebook	95
web.mac	91
folkloredelnorte	88
taringa	65
ensayistas	64
base.kb	54
adonde	48
razoncartografica	47
unc.edu	47
fr.wikipedia	43
Incaempire	42
elearning.uniroma1	42
www1.assumption.edu	42

Appendix 18 Referring Wikipedia pages

ADL referring Wikipedia pages (top 10) (#=number of sessions).

Wikipedia page	#
http://da.wikipedia.org/wiki/Georg_Brandes	59
http://en.wikipedia.org/wiki/Thumbelina	55
http://no.wikipedia.org/wiki/Ludvig_Holberg	55
http://en.wikipedia.org/wiki/The_Princess_and_the_Pea	49
http://da.wikipedia.org/wiki/Bl%C3%A5t%C3%A5rn_(K%C3%B8benhavns_Slot)	48
http://da.wikipedia.org/wiki/Adam_Oehlenschl%C3%A4ger	48
http://da.wikipedia.org/wiki/H.C._Andersen	47
http://da.wikipedia.org/wiki/J.P._Jacobsen	45
http://ja.wikipedia.org/wiki/%E8%A3%B8%E3%81%AE%E7%8E%8B%E6%A7%98	45
http://en.wikipedia.org/wiki/The_Emperor%27s_New_Clothes	43

KID referring Wikipedia pages (top 10) (#=number of sessions).

Wikipedia page	#
http://da.wikipedia.org/wiki/Weilbachs_Kunstnerleksikon	115
http://en.wikipedia.org/wiki/Vilhelm_Hammersh%C3%B8i	44
http://en.wikipedia.org/wiki/Wilhelm_Freddie	37
http://de.wikipedia.org/wiki/Weilbachs_K%C3%BCnstlerlexikon	30
http://da.wikipedia.org/wiki/Johannes_Larsen	25
http://da.wikipedia.org/wiki/Kunstindeks_Danmark	22
http://en.wikipedia.org/wiki/Paul_Gustave_Fischer	17
http://da.wikipedia.org/wiki/P.S._Kr%C3%B8yer	17
http://da.wikipedia.org/wiki/Fritz_Syberg	16
http://da.wikipedia.org/wiki/F.C._Lund_(maler)	16

Poma referring Wikipedia pages (top 10) (#=number of sessions).

Wikipedia page	#
http://es.wikipedia.org/wiki/Primer_Nueva_coronica_y_buen_gobierno	165
http://es.wikipedia.org/wiki/Felipe_Guam%C3%A1n_Poma_de_Ayala	164
http://en.wikipedia.org/wiki/Felipe_Guaman_Poma_de_Ayala	134
http://en.wikipedia.org/wiki/Nueva_Cr%C3%B3nica_y_Buen_Gobierno	88
http://en.wikipedia.org/wiki/Primer_Nueva_Cor%C3%B3nica_y_Buen_Gobierno	35
http://fr.wikipedia.org/wiki/Felipe_Guaman_Poma_de_Ayala	32
http://de.wikipedia.org/wiki/Waman_Puma_de_Ayala	31
http://en.wikipedia.org/wiki/Inca_Empire	27
http://en.wikipedia.org/wiki/Guaman_Poma	17
http://ru.wikipedia.org/wiki/[Guaman Poma] (in Cyrillic)	14

Appendix 19 Countries of origin

Users country of origin in ADL (share>1.0%).

Country	%
Denmark	78.6 %
United States	6.7 %
Norway	2.6 %
Germany	2.1 %
Sweden	1.2 %
China	1.2 %
Other	7.6 %

Users country of origin in KID (share>1.0%).

Country	%
Denmark	84.9 %
United States	2.7 %
Norway	2.0 %
China	1.8 %
Germany	1.4 %
Sweden	1.3 %
Other	4.7 %

Users country of origin in Poma (share>1.0%).

Country	%
Peru	35.3 %
United States	15.2 %
Mexico	11.1 %
Colombia	6.8 %
Spain	5.1 %
Argentina	4.1 %
Chile	3.9 %
Ecuador	3.1 %
Venezuela	1.6 %
United Kingdom	1.5 %
Bolivia	1.4 %
Denmark	1.1 %
Other	9.9 %

Appendix 20 Distribution of session paths

Distribution of session paths in ADL.

Session path	Number of sessions	Number of pages viewed	Avg.session length	Share of all sessions
N1	571	571	1,0	1,3%
N2	1147	11524	10,0	2,6%
N3	171	2840	16,6	0,4%
N4	2315	62327	26,9	5,2%
N5	583	2463	4,2	1,3%
I1	18716	18716	1,0	42,2%
I2	808	7056	8,7	1,8%
I3	3119	42464	13,6	7,0%
I4	966	23863	24,7	2,2%
I5	6681	23006	3,4	15,1%
O1	5509	5509	1,0	12,4%
O2	912	11828	13,0	2,1%
O3	138	2192	15,9	0,3%
O4	349	9824	28,1	0,8%
O5	2367	32443	13,7	5,3%
All	44352	256626	5,8	100,0%

Distribution of session paths in KID.

Session path	Number of sessions	Number of pages viewed	Avg. session length	Share of all sessions
N1	239	239	1,0	1,1%
N2	504	5271	10,5	2,2%
N3	2109	26822	12,7	9,3%
N4	4396	82072	18,7	19,4%
N5	238	1299	5,5	1,1%
I1	6508	6508	1,0	28,7%
I2	940	7757	8,3	4,1%
I3	1598	18463	11,6	7,1%
I4	1222	22380	18,3	5,4%
I5	2371	9234	3,9	10,5%
O1	576	576	1,0	2,5%
O2	275	2538	9,2	1,2%
O3	253	3299	13,0	1,1%
O4	651	12380	19,0	2,9%
O5	786	4234	5,4	3,5%
ALL	22666	203072	9,0	100,0%

Distribution of session paths in Poma.

Session path	Number of sessions	Number of pages viewed	Avg. session length	Share of all sessions
N1	9170	9170,00	1,0	30,0%
N4	4152	64263,00	15,5	13,6%
N5	408	1307,00	3,2	1,3%
O1	13787	13787,00	1,0	45,1%
O4	916	12041,00	13,1	3,0%
O5	2124	13597,00	6,4	7,0%
ALL	30557	114165,00	3,7	100,0%

Appendix 21 Distribution of session paths based on navigation strategy

ADL: Distribution of session paths based on navigation strategy.

Session path and navigation strategy	Number of sessions	Number of pages viewed	Avg. session length	Share of all sessions
N1 dir	8	8	1,0	0,02%
N1 link	150	150	1,0	0,34%
N1 se	413	413	1,0	0,93%
N2 dir	491	6186	12,6	1,11%
N2 link	177	1910	10,8	0,40%
N2 se	479	3428	7,2	1,08%
N3 dir	114	2187	19,2	0,26%
N3 link	20	273	13,7	0,05%
N3 se	37	380	10,3	0,08%
N4 dir	1407	42720	30,4	3,17%
N4 link	362	8717	24,1	0,82%
N4 se	546	10890	19,9	1,23%
N5 dir	125	778	6,2	0,28%
N5 link	82	409	5,0	0,18%
N5 se	376	1276	3,4	0,85%
I1 dir	33	33	1,0	0,07%
I1 link	321	321	1,0	0,72%
I1 se	18362	18362	1,0	41,40%
I2 dir	97	1140	11,8	0,22%
I2 link	54	554	10,3	0,12%
I2 se	657	5362	8,2	1,48%
I3 dir	245	5276	21,5	0,55%
I3 link	427	7561	17,7	0,96%
I3 se	2447	29627	12,1	5,52%
I4 dir	106	3102	29,3	0,24%
I4 link	147	4305	29,3	0,33%
I4 se	713	16456	23,1	1,61%
I5 dir	162	1055	6,5	0,37%
I5 link	172	990	5,8	0,39%
I5 se	6347	20961	3,3	14,31%
O1 dir	31	31	1,0	0,07%
O1 link	121	121	1,0	0,27%
O1 se	5357	5357	1,0	12,08%
O2 dir	147	2882	19,6	0,33%

Session path and navigation strategy	Number of sessions	Number of pages viewed	Avg. session length	Share of all sessions
O2 link	98	1605	16,4	0,22%
O2 se	667	7341	11,0	1,50%
O3 dir	46	969	21,1	0,10%
O3 link	14	402	28,7	0,03%
O3 se	78	821	10,5	0,18%
O4 dir	137	4592	33,5	0,31%
O4 link	40	1240	31,0	0,09%
O4 se	172	3992	23,2	0,39%
O5 dir	824	17217	20,9	1,86%
O5 link	329	5601	17,0	0,74%
O1 se	1214	9625	7,9	2,74%
TOTAL	44352	256626	5,8	100,00%

KID: Distribution of session paths based on navigation strategy.

Session path and navigation strategy	Number of sessions	Number of pages viewed	Avg. session length	Share of all sessions
N1 dir	26	26	1,0	0,1%
N1 link	108	108	1,0	0,5%
N1 se	105	105	1,0	0,5%
N2 dir	192	2111	11,0	0,8%
N2 link	114	1603	14,1	0,5%
N2 se	198	1557	7,9	0,9%
N3 dir	810	9827	12,1	3,6%
N3 link	646	8395	13,0	2,9%
N3 se	653	8600	13,2	2,9%
N4 dir	2001	35282	17,6	8,8%
N4 link	934	18392	19,7	4,1%
N4 se	1461	28398	19,4	6,4%
N5 dir	91	526	5,8	0,4%
N5 link	78	371	4,8	0,3%
N5 se	69	402	5,8	0,3%
I1 dir	220	220	1,0	1,0%
I1 link	1582	1582	1,0	7,0%
I1 se	4706	4706	1,0	20,8%
I2 dir	81	771	9,5	0,4%
I2 link	265	2162	8,2	1,2%
I2 se	594	4824	8,1	2,6%
I3 dir	62	786	12,7	0,3%
I3 link	518	7143	13,8	2,3%
I3 se	1018	10534	10,3	4,5%
I4 dir	77	1479	19,2	0,3%
I4 link	406	7498	18,5	1,8%
I4 se	739	13403	18,1	3,3%
I5 dir	22	115	5,2	0,1%
I5 link	882	3459	3,9	3,9%
I5 se	1467	5660	3,9	6,5%
O1 dir	3	3	1,0	0,0%
O1 link	275	275	1,0	1,2%
O1 se	298	298	1,0	1,3%
O2 dir	6	27	4,5	0,0%
O2 link	92	1115	12,1	0,4%
O2 se	177	1396	7,9	0,8%
O3 dir	10	67	6,7	0,0%
O3 link	200	2943	14,7	0,9%

Session path and navigation strategy	Number of sessions	Number of pages viewed	Avg. session length	Share of all sessions
O3 se	43	289	6,7	0,2%
O4 dir	12	240	20,0	0,1%
O4 link	558	10893	19,5	2,5%
O4 se	81	1247	15,4	0,4%
O5 dir	112	636	5,7	0,5%
O5 link	316	2309	7,3	1,4%
O1 se	358	1289	3,6	1,6%
ALL	22666	203072	9,0	100,0%

Poma: Distribution of session paths based on navigation strategy.

Session path and navigation strategy	Number of sessions	Number of pages viewed	Avg. session length	Share of all sessions
N1 dir	46	46	1,0	0,2%
N1 link	7590	7590	1,0	24,8%
N1 se	1534	1534	1,0	5,0%
N4 dir	996	18895	19,0	3,3%
N4 link	778	12571	16,2	2,5%
N4 se	2378	32797	13,8	7,8%
N5 dir	63	246	3,9	0,2%
N5 link	45	170	3,8	0,1%
N5 se	300	891	3,0	1,0%
O1 dir	46	46	1,0	0,2%
O1 link	400	400	1,0	1,3%
O1 se	13341	13341	1,0	43,7%
O4 dir	244	4411	18,1	0,8%
O4 link	82	1011	12,3	0,3%
O4 se	590	6619	11,2	1,9%
O5 dir	330	5801	17,6	1,1%
O5 link	79	373	4,7	0,3%
O5 se	1715	7423	4,3	5,6%
ALL	30557	114165,00	3,7	100,0%

Appendix 22 Navigation strategies, intentions and work contexts in survey

Frequency of answers to the question: "How did you reach this site?".

Answer	ADL #	ADL %	KID #	KID %	Poma #	Poma %
Direct via a bookmark	9	16.4%	38	14.8%	5	11.4%
Direct by typing in the URL	19	34.5%	35	13.7%	6	13.6%
Through links on the web (web pages, blogs, etc.)	10	18.2%	41	16.0%	13	29.5%
By using a search engine for a topical search (e.g. author or artists name as search terms)	10	18.2%	109	42.6%	14	31.8%
By using a search engine to find known site (e.g. parts of title or URL as search terms)	7	12.7%	33	12.9%	6	13.6%
Total	55	100%	256	100%	44	100%

Frequency of answers to the question: "In what context do you visit the web site?"

Answer	ADL #	ADL %	KID #	KID %	Poma #	Poma %
Hobby or leisure	24	43.6%	123	48.0%	12	27.3%
Work	14	25.5%	85	33.2%	5	11.4%
By coincidence	1	1.8%	7	2.7%	0	0%
School or study	14	25.5%	13	5.2%	22	50.0%
Other	2	3.6%	28	10.9%	5	11.4%
Total	55	100%	256	100%	44	100%

Frequency of answers to the question: "Why are you visiting this resource today?".

Answer	ADL #	ADL %	KID #	KID %	Poma #	Poma %
Exploring	19	34.5%	49	19.1%	10	22.7%
Learning	26	47.3%	181	70.7%	23	52.3%
Look up fact	18	32.7%	158	61.7%	5	11.4%
Curious	11	20.0%	39	15.2%	5	11.4%
Other	13	26.3%	44	17.2%	8	18.2%
Total	*87*		*471*		*51*	

Appendix 23 Crosstabulations on survey data

ADL: Crosstabulation of Task context and Navigation strategy.

	Direct via a bookmark		Direct by typing in the url		Through links on the web		By using a search engine for a topical search		By using a search engine to find known site		Total	
Hobby or leisure	5	55.6%	8	42.1%	3	30.0%	6	60.0%	2	28.6%	24	43.6%
Work	1	11.1%	8	42.1%	2	20.0%	1	10.0%	2	28.6%	14	25.5%
By coincidence	0	0%	0	0%	0	0%	1	10.0%	0	0%	1	1.8%
School or study	2	22.2%	3	15.8%	4	40.0%	2	20.0%	3	42.9%	14	25.5%
Other	1	11.1%	0	0%	1	10.0%	0	0%	0	0%	2	3.6%
All	9	100%	19	100%	10	100%	10	100%	7	100%	55	100%

(no statistically significant chi-relationship)

ADL: Crosstabulation of Task context and Gender (n=55).

	Female #	**Female %**	**Male #**	**Male #**	**Total #**	**Total %**
Hobby or leisure	7	31.8%	17	51.5%	24	43.6%
Work	7	31.8%	7	21.2%	14	25.5%
By coincidence	0	0%	1	3.0%	1	1.8%
School or study	8	36.4%	6	18.2%	14	25.5%
Other	0	0%	2	6.1%	2	3.6%
	22	100%	33	100%	55	100%

(no statistically significant chi-relationship)

KID: Crosstabulation of Task context and Gender.

	Female	Male	Total
Hobby or leisure	35.1%	58.0%	48.0%
Work	43.0%	25.4%	33.2%
By coincidence	4.4%	1.4%	2.7%
School or study	7.9%	2.8%	5.1%
Other	9.6%	12.0%	10.9%
	100%	100%	100%

(χ^2=18.676, df=4, p<0.01)

KID: Crosstabulation of Task context and Age groups.

	19-30	31-45	46-65	66-	Total
Hobby or leisure	26.7%	26.1%	51.2%	61.4%	48.0%
Work	53.3%	67.4%	28.0%	15.7%	33.2%
By coincidence		2.2%	3.2%	2.9%	2.7%
School or study	20.0%	4.3%	4.8%	2.9%	5.1%
Other			12.8%	17.1%	10.9%
Total	100%	100%	100%	100%	100%

(χ^2=51.826, df=12, p<0.01)

KID: Crosstabulation of Task context and Navigation strategy.

	Direct via a bookmark	Direct by typing in the url	Through links on the web	By using a search engine for a topical search	By using a search engine to find known site	Total
Hobby or leisure	31.6%	17.1%	61.0%	56.9%	54.5%	48.0%
Work	47.4%	80.0%	24.4%	17.4%	30.3%	33.2%
By coincidence	0%	0%	4.9%	3.7%	3.0%	2.7%
School or study	2.6%	2.9%	4.9%	7.3%	3.0%	5.1%
Other	18.4%	0%	4.9%	14.7%	9.1%	10.9%
All	100%	100%	100%	100%	100%	100%

(χ^2=61.032, df=16, p<0.01)

Poma: Crosstabulation of Task context and Navigation strategy.

	Direct via a bookmark		Direct by typing in the url		Through links on the web		By using a search engine for a topical search		By using a search engine to find known site		Total	
Hobby or leisure	0	0%	3	50.0%	3	23.1%	4	26.8%	2	33.3%	12	27.3%
Work	1	20.0%	0	0%	2	15.4%	2	14.3%	0	0%	5	11.4%
By coincidence	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
School or study visit	4	80.0%	1	16.7%	7	53.8%	7	50.0%	3	50.0%	22	50.0%
Other	0	0%	2	33.3%	1	7.7%	1	7.1%	1	16.7%	5	11.4%
Total	5	100%	6	100%	13	100%	14	100%	6	100%	44	100%

(no statistically significant chi-relationship)

Poma: Crosstabulation of Task context and Gender (n=44).

	Female #	Female %	Male #	Male %	Total #	Total %
Hobby or leisure	4	20%	8	33%	12	27%
Work	2	10%	3	13%	5	11%
By coincidence	0	0%	0	0%	0	0%
School or study	12	60%	10	42%	22	50%
Other	2	10%	3	13%	5	11%
Total	20	100%	24	100%	44	100%

(no statistically significant chi-relationship)

Appendix 24 Comparison logs and survey

Distribution of navigation strategies in logs and survey.

	ADL Log n=44352		**ADL Survey n=55**		**KID Log n=22667**		**KID Survey n=256**		**Poma Log n=30557**		**Poma Survey n=44**	
Direct	3973	9.0%	28	51%	3750	17%	73	29%	1725	5.6%	23	52.3%
Link	2514	5.7%	10	18%	6941	31%	41	16%	8974	29.4%	2	4.5%
Search Engine	37865	85.4%	17	31%	11976	53%	142	56%	19858	65.0%	19	43.2%
Total	*44352*	*100%*	*55*	*100%*	*22667*	*100%*	*256*	*100%*	*30557*	*100%*	*44*	*100%*

Users country of origin in logs and survey in all three resources.

Country	ADL Log (n=44352)	ADL Survey (n=55)	KID Log (n=22667)	KID Survey (n=256)	Poma Log (n=30557)	Poma Survey (n=44)
Denmark	78.6%	76.4%	89.2%	89.5%	1.1%	6.8%
Other country	21.4%	23.6%	10.8%	10.5%	98.9%	93.2%
Total	*100%*	*100%*	*100%*	*100%*	*100%*	*100%*